

eScholarship@UMassChan

Characterizing Colibactin Toxicity and the Resulting Cellular Response to DNA Damage in Mammalian and Bacterial Systems

Item Type	Doctoral Dissertation
Authors	Lowry, Emily
DOI	10.13028/33g6-3d15
Publisher	UMass Chan Medical School
Rights	Copyright © 2024 Emily Lowry
Download date	2026-04-10 15:35:54
Item License	https://creativecommons.org/licenses/by-nc/4.0/
Link to Item	https://hdl.handle.net/20.500.14038/53830

CHARACTERIZING COLIBACTIN
TOXICITY AND THE RESULTING
CELLULAR RESPONSE TO DNA DAMAGE
IN MAMMALIAN AND BACTERIAL
SYSTEMS

A Dissertation Presented

By

EMILY LOWRY

Submitted to the Faculty of the
Morningside Graduate School of Biomedical Sciences at UMass Chan Medical
School

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

CHARACTERIZING COLIBACTIN TOXICITY AND THE RESULTING
CELLULAR RESPONSE TO DNA DAMAGE IN MAMMALIAN AND
BACTERIAL SYSTEMS

A Dissertation Presented By

EMILY LOWRY

This work was undertaken in the Morningside Graduate School of Biomedical
Sciences

Cancer Biology

Under the mentorship of

Amir Mitchell, Thesis Advisor

Michael Lee, Member of Committee

Jennifer Benanti, Member of Committee

Josué Flores Kim, Member of Committee

Babak Momeni, External Member of Committee

Elizabeth Shank, Chair of Committee

Mary Ellen Lane, Ph.D.,

Dean of the Morningside Graduate School of Biomedical Science

September 3rd, 2024

Dedication

This thesis is dedicated to the women who have come before me, bravely laying the groundwork for future generations to pursue higher education.

Acknowledgements

I first need to thank **Amir Mitchell** for mentoring me throughout my PhD and giving me the opportunity to work in his lab and learn from him. Amir is incredibly intelligent and an amazing story-teller and writer. I am grateful that I have had exposure to these skills and hopefully show growth in presenting and writing from Amir. I also need to mention the collaborative and welcoming lab environment that speaks volumes of the group dynamic that Amir fosters. When I first joined the lab, Amir saw potential that I didn't know I had. He encouraged me to dive headfirst into new assays and to conduct the first CRISPR screen in the lab. Reflecting back on myself as a first-year grad student, I can see how much I have grown and I thank Amir for giving me the chance and space to do so in his lab.

Thank you to my TRAC and DEC members throughout my time at UMass. Thank you, **Beth Shank, Mike Lee, Jenny Benanti, Josué Flores Kim, Sharon Cantor** for your guidance at each TRAC meeting. Thank you **Babak Momeni** for serving as my external examiner, as well.

I also need to acknowledge the mentors and teachers who helped get me here. I was never the most confident student, but there were a few teachers over the years to noticed my interest in science and went the extra step to encourage it. Those little actions of giving me a few extra minutes to chat about the ocean or tell me that my science fair project was amazing (even if my judges trashed it) made all the difference in me continuing to pursue a career in science. Thank you to **Professor Stoehr** who took me on as a student researcher and even encouraged

me to apply for a research grant in undergrad. Thank you, **Michal** for hiring me and giving me the amazing opportunity to work with cell lines, genetic screens, and NK cells. I learned so much more than I thought I could from a technician position and you have continued to be a supportive mentor during my PhD.

To the Mitchell Lab: thank you! **Serkan**, you have always been such a positive force in the lab and so supportive and helpful. From teaching me everything I know about bacterial cloning to yelling at me to just get my coffee and get my work done (I hope you still have that sticker!). You're the best senior grad student anyone could hope to have. **Brittany**, the queen of pre-experiment organizing and planning. Thank you for being a model student and also for all of your dog knowledge. **Mariana**, my cohort buddy in the lab! Thanks for encouraging me to rotate with Amir in the first place and also always helping me with my code. Thank you also for pushing me to start salsa dancing because I ended up loving it. **Sydney**, my original mammalian cell and colibactin buddy. Your perseverance is inspiring. You tackle everything with a positive attitude and such a thorough and critical approach that is going to help you succeed in your career. And thanks for being a fellow Swiftie and trash TV fan (woohoo bachelor/bachelorette). **Carmen**, my almost birthday buddy and fellow Massachusetts native (and rotation mentee!). You are also so positive and amazing at tackling huge experiments without batting an eye. Thank you for always telling us about free food opportunities and leading the charge to go get free food so we don't have to do it alone. **Michelle**, the newest member of the lab and one of my rotation mentees. You are so positive and thoughtful and I can't

wait to watch you grow in your PhD and as a future physician-scientist. Thanks for also being a fellow trash TV fan and analyzing the bachelor/bachelorette. I am so thankful that the Mitchell lab is a welcoming space to everyone and that we have made the effort to be inclusive and hang out (like our famous movie nights!).

I also need to thank many other labs and members of DSB. I felt a little out of place when I first joined the Mitchell lab since I had mostly done my rotation unofficially during the covid shutdown, so no one in the department really knew who I was. It took a while since we were on shift work and some people worked remotely, but I eventually got to know everyone and felt more at home. Many alumni members of the Walhout lab were my lunch buddies early on. **Brent, Yonguk, Olga, Nana, Sushila, Cedric, Gabby**, thank you for just being there so I didn't have to eat lunch alone. **Brent**, my fellow Harrington, Clarke, LHS, and now UMass alum, I am so glad that I met you here and became friends and lunch buddies. You are so weird and it's great and I am so happy that we are hiking buddies too. **Yonguk**, you are also so weird and it's great. Thank you for being a great friend, lunch buddy, and life mentor. **Sunny**, thank you for being so friendly right off the bat! **Vinu**, thank you for teaching me so much about *E. coli* and ALWAYS answering my questions. **Sunil**, thank you for also teaching me about *E. coli* and answering my questions when Vinu wasn't around, and thank you for being a valued honorary member of the Mitchell lab for movie nights. Thanks to the **Shank lab** for always providing some good laughs between our lab bays and helping me with general bacteria experiment questions. Thanks

to the **Lee lab**, and especially **Megan**, for sharing the tissue culture room and answering all my random cell and CRISPR screen questions.

To my cohort: thanks for keeping things fun and light, even when the world came to an abrupt halt when covid hit.

Judy and **Milky**, thank you for being the best roommates and friends and support network. We are so lucky that we agreed to live together, barely knowing each other, and that it worked out so well. Also **Kevin**, who joined the roomie crew a few months in, thanks for being part of such a nice and friendly and supportive group. We were the best covid pod squad! Thank you guys for putting up with me making us celebrate holidays together with elaborate meals. And my constant singing and “hey google, can you please play ‘chill hits’ on spotify”. And supporting my crazy Taylor Swift obsession. And all getting dogs so our dogs can be best friends too.

My darling **Daisy**. You are the best thing that has ever happened to me. You are so weird and cuddly and funny and I love you so much my little fluffball. You have been the best snuggler when I’ve worked from home, even if you sometimes tried to climb onto my laptop. Even on bad days, coming home to your happy face and wagging tail made me instantly feel better. You bring me so much comfort and happiness and unconditional love. You’re my soul dog through and through.

I have grown in my independence and confidence in trying new things and seeking out strangers to do things I know I love. Thank you to the AMC Worcester 20s and 30s chapter for giving me a chance to get out hiking with new

people, for camping with me, and for giving me a bunkhouse to stay in with a bunch of strangers to experience an eclipse in totality. Thank you Plantation 2 Peaks for introducing me to hikers at UMass, including **Josias**. I tried winter hiking for the first time with you and now have hiking buddies to bag peaks with! Shout out to Salsa Worcester for providing a fun and safe environment to learn salsa dancing. I love that every week I can dance and have fun with new and old faces and even feel comfortable going social dancing in Boston. Another special thanks to the ladies who take classes at the UMass gym! I always liked working out, but the camaraderie and support from this group makes me look forward to going to the gym and challenging myself in my workouts. Thank you, **Raquel, Jess, Jennifer, Ebru, Cle, Carmen, Milky, Monika**, and many others!

I am fortunate to have maintained several friendships from college. **Allison**, thanks for listening to my PhD stress rants and coming to Worcester often to have dinner with me. **Abby**, thank you for also listening to my PhD stress rants and always being down to get Los Andes with me. Abby and **Kate**, thanks for the annual girls weekends featuring a lot of seafood.

The galz! Thanks **Monika, Judy, Milky**, and **Stephanie** for creating a space for us to all vent, gossip, and drink cocktails.

Thank you to my family for putting up with my insane nerdiness and random science facts for decades. At least I'm doing something with my nerdiness! Thanks for always giving me a space to take a break when we are all together. And thanks in advance for not hating me when I brag about being the only doctor in the family because I think I've earned the right to rub it in a little.

Finally, thank you Taylor Swift, the queen of my heart. Your music got me through the highs and lows of grad school and was the soundtrack to my late nights in the lab (sorry to anyone who may have heard me singing/seen me dancing), meal prepping on Sundays, and any long drives.

I am sure I have missed important acknowledgements because so many people had an impact on my life these last five years, so I want to dedicate an additional thank you to anyone I may have left out – it's not intentional, my brain is just dead.

Abstract

The bacterial toxin colibactin, produced primarily by the B2 phylogroup of *Escherichia coli*, crosslinks DNA and can promote colon cancer in human hosts, where it has been extensively studied. A systematic approach to identify the DNA damage response to colibactin-induced toxicity has yet to be applied and colibactin toxicity in bacteria remains underexplored. Using a genome-wide CRISPR screen in colon cancer cells, I found that colibactin activates most DNA repair pathways with key roles for Fanconi anemia/interstrand crosslink repair and fork quality control pathways. I also conducted a genome-wide loss-of-function screen in *E. coli* that identified a key role for homologous recombination in repairing colibactin-induced damage. I determined that colibactin induces a mutational pattern in *E. coli* in A/T rich regions, as it does in colon cells, but that the resulting mutational signature differs in *E. coli*. I then predicted that long-term colibactin exposure will culminate in a genomic bias based on this mutational signature, which may be detected in colibactin-producing bacteria. I tested this prediction by analyzing thousands of *E. coli* genomes and found that colibactin-producing strains show skewness in trinucleotide composition. Finally, I used a sensitive DNA damage reporter assay to find that cell-cell contact is not required in bacteria as was previously suggested for both bacteria and mammalian cells, and that this needs to be reevaluated in mammalian cells. Taken together, this work revealed the DNA damage response to colibactin-induced damage in both colon and bacteria cells, a bacteria-specific mutation pattern, and that cell-cell contact is not required in bacteria.

Table of Contents

DEDICATION	III
ACKNOWLEDGEMENTS	IV
ABSTRACT.....	X
TABLE OF CONTENTS	XI
LIST OF TABLES.....	XIV
LIST OF FIGURES.....	XV
LIST OF COPYRIGHTED MATERIALS	XVI
LIST OF ABBREVIATIONS	XVII
LIST OF MULTIMEDIA FILES.....	XVIII
CHAPTER 1: INTRODUCTION.....	1
1.1 THE HUMAN MICROBIOME AND ITS ROLE IN DISEASE	1
1.1.1 <i>Overview of the human microbiome</i>	1
1.1.2 <i>Associations of the gut microbiome with disease</i>	2
1.1.3 <i>Tumor-promoting bacteria – indirect causes</i>	3
1.1.4 <i>Tumor-promoting bacteria – direct causes</i>	5
1.2 BACTERIA IN THE TUMOR MICROBIOME	7
1.3 COLIBACTIN, A TUMORIGENIC TOXIN	8
1.3.1 <i>Overview of colibactin</i>	8
1.3.2 <i>Colibactin and its role in cancer</i>	11
1.4 BACTERIA-BACTERIA COMPETITION.....	13
1.4.1 <i>Overview of microbial competition</i>	13
1.4.2 <i>Nutrient uptake as an approach to competition</i>	13
1.4.3 <i>Bacteria warfare</i>	15
1.4.4 <i>Colibactin as mechanism of interference competition</i>	16
1.4.5 <i>Colibactin toxicity to neighboring bacteria</i>	17
1.5 DNA DAMAGE REPAIR PATHWAYS IN HUMANS AND <i>E. COLI</i>	19
1.5.1 <i>Overview of DNA damage repair</i>	19
1.5.2 <i>Base excision repair</i>	20
1.5.3 <i>Nucleotide excision repair</i>	20
1.5.4 <i>Mismatch repair</i>	21
1.5.5 <i>Non-homologous end joining</i>	22
1.5.6 <i>Homologous recombination</i>	23
1.5.7 <i>Other repair pathways</i>	23
1.6 MUTATIONAL BIASES	25
1.6.1 <i>Mutation bias in human cells from DNA repair</i>	25
1.6.2 <i>Mutation bias in E. coli</i>	26
1.6.3 <i>Damage-induced mutational signatures</i>	27
1.7 SCOPE OF THIS THESIS	28
CHAPTER 2: SYSTEMATIC MAPPING OF THE DNA DAMAGE RESPONSE TO COLIBACTIN EXPOSURE IN COLON CANCER.....	31
2.1 ABSTRACT	31
2.2 INTRODUCTION.....	32
2.3 RESULTS.....	34
2.3.1 <i>Colibactin induces dose-dependent DNA damage</i>	34
2.3.2 <i>Colibactin induces cell cycle arrest in a dose-dependent manner</i>	36
2.3.3 <i>Genetic screen uncovers the DNA damage response to colibactin</i>	40
2.4 DISCUSSION.....	45

2.5	METHODS.....	48
2.5.1	<i>Cell lines and media</i>	49
2.5.2	<i>Cloning deletion strains</i>	49
2.5.3	<i>Co-culture assays in 96-well format</i>	49
2.5.4	<i>Immunofluorescence staining</i>	50
2.5.5	<i>Immunofluorescence microscopy and analysis</i>	51
2.5.6	<i>Proliferation assays</i>	52
2.5.7	<i>TKO library generation</i>	52
2.5.8	<i>CRISPR screen</i>	53
2.5.9	<i>CRISPR screen sequencing</i>	54
2.5.10	<i>CRISPR screen analysis</i>	56
2.5.11	<i>Data and code availability</i>	56
CHAPTER 3: COLIBACTIN LEADS TO A BACTERIA-SPECIFIC MUTATION PATTERN AND SELF-INFLICTED DNA DAMAGE		57
3.1	PREFACE.....	57
3.2	ABSTRACT.....	57
3.3	INTRODUCTION.....	58
3.4	RESULTS.....	62
3.4.1	<i>Colibactin reduces viability of co-cultured E. coli</i>	62
3.4.2	<i>Genetic screen uncovers genes mitigating colibactin toxicity</i>	63
3.4.3	<i>Colibactin induces a specific mutational signature</i>	69
3.4.4	<i>Self-inflicted damage in colibactin-producing E. coli</i>	73
3.5	DISCUSSION.....	78
3.6	METHODS.....	82
3.6.1	<i>Media and growth conditions</i>	83
3.6.2	<i>Cloning deletion strains</i>	83
3.6.3	<i>Monitoring colibactin impact on viability of co-cultured cells</i>	83
3.6.4	<i>Genome-wide loss-of-function genetic screen</i>	84
3.6.5	<i>Analysis of genetic screen results</i>	87
3.6.6	<i>Targeted validation genetic screen</i>	87
3.6.7	<i>Mutation accumulation</i>	89
3.6.8	<i>DNA damage fluorescent reporter</i>	91
3.6.9	<i>Self-inflicted damage microscopy</i>	91
3.6.10	<i>Bioinformatics analysis of colibactin-linked trinucleotide skew</i>	92
3.6.11	<i>Data Availability</i>	93
CHAPTER 4: COLIBACTIN-INDUCED DAMAGE IN BACTERIA IS CELL CONTACT INDEPENDENT		94
4.1	PREFACE.....	94
4.2	ABSTRACT.....	94
4.3	IMPORTANCE.....	95
4.4	INTRODUCTION.....	95
4.5	RESULTS.....	99
4.5.1	<i>Colibactin induces DNA damage in E. coli</i>	99
4.5.2	<i>Colibactin induces DNA damage in distant cells</i>	102
4.5.3	<i>Colibactin-induced DNA damage is independent of cell contact</i>	105
4.6	DISCUSSION.....	113
4.7	METHODS.....	116
4.7.1	<i>Media and growth conditions</i>	116
4.7.2	<i>Cloning deletion strains</i>	117
4.7.3	<i>Fluorescent reporter plasmids</i>	117
4.7.4	<i>Monitoring DNA damage reporter with flow cytometry</i>	118
4.7.5	<i>Monitoring DNA damage reporter by microscopy in a lawn co-culture</i>	119
4.7.6	<i>Monitoring DNA damage reporter by microscopy in single colonies</i>	121
4.6.7	<i>Colony cross contamination test</i>	122

4.7.7	Code availability.....	123
CHAPTER 5:	DISCUSSION	124
5.1	SUMMARY	124
5.2	LIMITATIONS OF THIS WORK.....	125
5.3	IMPLICATIONS OF THIS WORK	128
5.3.1	<i>Potential for host-specific colibactin therapeutics</i>	<i>128</i>
5.3.2	<i>Colibactin-induced DNA damage as a biomarker.....</i>	<i>130</i>
5.3.3	<i>Bacterial metabolites in cancer.....</i>	<i>132</i>
5.4	CONCLUSIONS	134
CHAPTER 6:	APPENDIX I: RAPID SIGNALING REACTIVATION AFTER TARGETED BRAF INHIBITION PREDICTS THE PROLIFERATION OF INDIVIDUAL MELANOMA CELLS FROM AN ISOGENIC POPULATION	136
6.1	PREFACE.....	136
6.2	ABSTRACT	137
6.3	INTRODUCTION.....	138
6.4	RESULTS.....	145
6.4.1	<i>Bulk measurement of signaling inhibition and reactivation.....</i>	<i>145</i>
6.4.2	<i>Bulk measurement of growth arrest and recovery.....</i>	<i>147</i>
6.4.3	<i>Single-cell measurement of signaling inhibition and reactivation</i>	<i>150</i>
6.4.4	<i>Single-cell measurement of growth recovery</i>	<i>158</i>
6.5	DISCUSSION.....	163
6.6	METHODS.....	166
6.6.1	<i>Cell-lines and media and antibodies</i>	<i>166</i>
6.6.2	<i>Live-cell fluorescent reporters.....</i>	<i>166</i>
6.6.3	<i>72-h dose response (ICW)</i>	<i>167</i>
6.6.4	<i>pERK peak inhibition and recovery (ICW).....</i>	<i>167</i>
6.6.5	<i>In-cell western</i>	<i>168</i>
6.6.6	<i>Time-lapse microscopy and automated image analysis</i>	<i>168</i>
BIBLIOGRAPHY		171

List of Tables

Table 1.1 pks island genes	9
Table 1.2 Repair pathways activated by types of DNA damage	20
Table 2.1 Bacteria strains used in this study	48
Table 2.2 Cell lines used in this study	48
Table 2.3 One-step PCR protocol	55
Table 3.1 Bacteria strains used in this study	82
Table 4.1 Strain purity in non-contacting colonies	111
Table 4.2 Bacteria strains used in this study	116

List of Figures

Figure 1.1 Colibactin is produced and exported by the pks island.	10
Figure 1.2 Bacteria take varying approaches to iron uptake.	15
Figure 2.1 Colibactin induces DNA damage in a dose-dependent manner.	35
Figure 2.2 Colibactin, and not the presence of bacteria, induces cell cycle arrest.	38
Figure 2.3 Genetic screen validates involvement of DNA repair genes in colibactin response.	41
Figure 2.4 Colibactin-induced damage activates checkpoints and multiple DNA repair pathways.	44
Figure 3.1 Genetic screen reveals key role of homologous recombination in colibactin response.	65
Figure 3.2 Pathway enrichment and targeted validation screen identify role of homologous recombination in colibactin response.	67
Figure 3.3 Colibactin induces a bacteria-specific mutational signature.	71
Figure 3.4 Colibactin inflicts self-damage.	75
Figure 3.5 Colibactin inflicts more self-damage in the absence of ClbS.	76
Figure 4.1 Temporal dynamics of colibactin-induced DNA damage.	100
Figure 4.2 Validation of the recA DNA damage reporter.	101
Figure 4.3 Colibactin induces DNA damage in distant cells.	103
Figure 4.4 Colibactin-induced DNA damage is cell-contact independent.	107
Figure 4.5 Colibactin-induced DNA damage decay similarly with distance in contacting and non-contacting colonies.	110
Figure 4.6 Validation of strain purity in the edge of non-contacting colonies	113
Figure 6.1 Reactivation of ERK signaling after vemurafenib treatment in bulk populations.	141
Figure 6.2 Transient growth arrest after vemurafenib treatment in bulk populations.	148
Figure 6.3 Growth recovery depends on drug concentrations and media replenishment protocol.	150
Figure 6.4 The molecular mechanism underlying the kinase translocation reporter (KTR) previously developed by the Covert lab and used in this study.	152
Figure 6.5 Single-cell measurements of signaling reveal homogenous inhibition yet heterogeneous signaling reactivation in WM983B cells.	153
Figure 6.6 Biological replication of the experiment presented in Figure 6.5.	157
Figure 6.7 Single-cell measurements connect early signaling recovery with long term drug resistance.	161
Figure 6.8 Biological replicates used in the experiment presented in Figure 6.7.	162

List of copyrighted materials

Figure 1.1A chemical structures of colibactin precursors is from the following publication:

Faïs T, Delmas J, Barnich N, Bonnet R, Dalmasso G. 2018. Colibactin: More Than a New Bacterial Toxin. *Toxins (Basel)* **10**: 151.

Figure 1.1A chemical structure of colibactin is from the following publication:

Dougherty MW, Jobin C. 2021. Shining a Light on Colibactin Biology. *Toxins (Basel)* **13**: 346. <https://pubmed.ncbi.nlm.nih.gov/34065799>.

The work in Chapter 6 (appendix I) is adapted from the following publication:

Khoshkenar P, Lowry E, Mitchell A. 2021. Rapid signaling reactivation after targeted BRAF inhibition predicts the proliferation of individual melanoma cells from an isogenic population. *Sci Rep* **11**: 15473. <https://doi.org/10.1038/s41598-021-94941-8>.

List of Abbreviations

A-EJ	Alternative end joining
AP	Apurinic/aprimidinic
BER	Base excision repair
BFT	<i>Bacteroides fragilis</i> toxin
CDT	cytolethal distending toxin
CRC	Colorectal cancer
DSB	Double strand break
EcN	<i>E. coli</i> Nissle 1917
FA	Fanconi Anemia
GGR	Global genome repair
HMP	Human Microbiome Project
HR	homologous recombination
HR	Homologous recombination
ICL	Interstrand crosslink
MMR	Mismatch repair
NER	Nucleotide excision repair
NHEJ	non-homologous end-joining
NHEJ	Non-homologous end joining
NMDA	<i>N</i> -myristoyl-D-Asn
NRPS	Non-ribosomal peptide synthetase
PKS	Polyketide synthase
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
sBLISS	in-suspension breaks labeling in situ and sequencing
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
T6SS	Type 6 secretion system
TCR	Transcription-coupled repair
TKO	Toronto knockout library
TLS	Translesion synthesis
UV	ultraviolet

List of Multimedia Files

Supplementary File 1 – colibactin CRISPR screen results (fold-change, empirical p-values)

Supplementary File 2 – colibactin CRISPR screen gene set enrichment analysis results

Supplementary File 3 – colibactin CRISPR screen DNA damage Fisher's exact test results

Supplementary File 4 – colibactin genetic screen in bacteria results (fold-change, p-values)

Supplementary File 5 – colibactin genetic screen in bacteria pathway enrichment

Supplementary File 6 – colibactin targeted screen in bacteria results (fold-change, p-values)

Supplementary File 7 – mutation accumulation results

Supplementary File 8 – *E. coli* genomes colibactin skew results

Chapter 1: Introduction

1.1 The human microbiome and its role in disease

1.1.1 *Overview of the human microbiome*

The human body contains an estimated 38 trillion bacteria cells, which compares to the approximately 30 trillion cells of the human body (Sender et al. 2016). The prevalence and abundance of microbes across the human body prompted the NIH to fund the Human Microbiome Project (HMP), which originally sought to characterize the human microbiome and eventually make connections to health and disease. The HMP sampled 15-18 sites on the human body from 242 adults and used 16S ribosomal sequencing, and whole-genome sequencing on a subset of samples, to identify the bacterial species present in the human microbiome (Methé et al. 2012). This approach also identified novel genes and established more than 3,000 reference genomes. The data collected from the HMP suggests that individuals harbor 500-1,000 different species of bacteria, while the number of species found across humans is ~10,000 (Turnbaugh et al. 2007; Gilbert et al. 2018). Despite the diversity of bacteria in the human body, the majority of species represent the *Firmicutes* and *Bacteroidota* phyla (Turnbaugh et al. 2007). Trends in species composition are associated with healthier gut microbiomes, while disruption of the microbiome, known as dysbiosis, can be associated with certain disease risks.

1.1.2 *Associations of the gut microbiome with disease*

The second phase of the HMP aimed to associate microbes or certain microbial genes with human disease. A major challenge to this goal is the diversity of the human microbiome between individuals. This diversity makes it challenging to identify species associated with health and disease, thus many connections are made at the phylum level. In general, decreases in microbial diversity in an individual are associated with various health issues such as adiposity, insulin resistance, and inflammation (Fan and Pedersen 2021; Hou et al. 2022). In fact, the gut microbiome is extensively studied for its associations with diabetes and obesity. A crucial study found that transferring the gut microbiome of an obese mouse to a germ-free mouse led to an increase in body weight of the recipient mouse, suggesting that the composition of the gut microbiome contributes to energy absorption of the host (Turnbaugh et al. 2006). Another study found that just nine species could inform if an individual is likely to be obese and likely to be at risk of diabetes. This conclusion was based on the presence of certain species in lean individuals and lack of these species in obese individuals, and vice versa (Le Chatelier et al. 2013). Of these nine species, those present only in lean individuals tend to be anti-inflammatory, while the bacteria present only in obese individuals tend to be pro-inflammatory.

Alternatively, respiratory diseases, such as asthma and chronic obstructive pulmonary disease, are associated with an increase in diversity and bacterial load in lungs (Hou et al. 2022; Hilty et al. 2010). The bacterial composition of the lungs can further change during exacerbations of respiratory diseases (Hou et al.

2022). The association of bacteria with respiratory diseases likely relates to inflammatory bacteria. *Proteobacteria* can aggravate airway diseases by promoting increased inflammatory cytokine expression and neutrophil infiltration (Huang et al. 2011; Alnahas et al. 2017).

Inflammatory bowel disease, which includes Crohn's disease and ulcerative colitis, is caused by many factors, but typically involves dysfunction of the immune system and disruption of the intestinal barrier. The genetics underlying irritable bowel disease can lead to immune dysregulation where immune cells become sensitive to intestinal antigens, triggering a cascade of pro-inflammatory responses, which can further contribute to disruptions to the intestinal barrier (Kaistha and Levine 2014). This immune dysregulation can also lead to a strong response in T cells to clear commensal bacteria (Kaistha and Levine 2014). Clearing of commensal bacteria by T cells creates a niche for bacterial pathogens and the opportunity for them to invade the intestinal barrier. Pathogens can more easily cause infections when the intestinal barrier is disrupted (McGuckin et al. 2009). This cycle of inflammation and infection can lead to lasting damage in the intestines.

1.1.3 *Tumor-promoting bacteria – indirect causes*

The general inflammatory nature of certain microbes has led to their association with tumor formation. One category of these bacteria is called “complicit” microbes, as they are carcinogenic, but are not sufficient to promote tumorigenesis on their own (Sepich-Poore et al. 2024). Commensal lung bacteria

can act as complicit microbes by activating T cells, resulting in tumor-promoting cytokine release. The inflammation caused by these bacteria in combination with a Kras mutation and loss of p53 function leads to lung cancer. Interestingly, in germ-free or antibiotic-treated mouse models lacking a lung microbiome, these mutations are insufficient to produce tumors (Jin et al. 2019). Additionally, the abundance of primary bile acid-producing bacteria, relative to bile acid metabolizing bacteria, in the colon increases the expression of CXCL16 on liver cells. This leads to activation of natural killer T cells and liver tumor growth. This tumorigenic effect is mitigated by supplementation of secondary bile acids or of bile acid metabolizing bacteria (Ma et al. 2018). While these two examples have been more clearly studied and outlined, many additional mechanisms through which bacteria may promote tumor formation are under studied. This is partly due to the rigorous experiments required to make these connections, and sometimes the low frequency of clinical prevalence.

There are also bacteria that can amplify tumorigenesis, but are unable to promote tumorigenesis on their own. These tumorigenic bacteria include *Fusobacterium nucleatum* and *Salmonella* spp. *F. nucleatum* produces the FadA adhesin that mediates epithelial cell invasion. FadA is then able to bind to E-cadherin, activating β -catenin signaling (Rubinstein et al. 2013). This ultimately results in an increased expression of transcription factors, oncogenes, and inflammatory genes, and also stimulates the growth of colorectal cancer cells (Rubinstein et al. 2013). Microbial gene sequencing from tumor samples has found over-representation of *fadA* compared to normal tissue, suggesting that the

fadA gene could be used as a diagnostic marker (Rubinstein et al. 2013).

Salmonella spp. produce the virulence factor AvrA, which inhibits NF- κ B signaling downstream of IKK, resulting in pro-inflammatory signaling (Collier-Hyams et al. 2002). In a cancer mouse model, AvrA increased β -catenin signaling and the number of tumors in the colon, but was insufficient to promote tumorigenesis alone (Lu et al. 2014).

1.1.4 *Tumor-promoting bacteria – direct causes*

While there are many avenues through which bacteria may contribute, but not cause, cancer, there are several known bacterial products that alone can induce mutations and lead to tumorigenesis. The first of these is cytolethal distending toxin (CDT), which is produced by gram negative bacteria, including *Escherichia coli*, *Campylobacter jejuni*, *Helicobacter* spp, and *Shigella* spp (Cortes-Bratti et al. 2001). CDT has DNAase activity and induces DNA double strand breaks (He et al. 2019). CDT-induced DNA damage is repaired by homologous recombination (HR), non-homologous end-joining (NHEJ) and possibly by the Fanconi Anemia (FA) pathway in the case of CDT-induced single strand breaks at low doses (Bezine et al. 2016). Cells exposed to CDT show replicative stress resulting in genomic instability that is passed onto daughter cells (Tremblay et al. 2021). This genomic instability may be the key factor in CDT carcinogenicity. In fact, Germ-free APC^{min/+} mice colonized with CDT-producing bacteria developed more tumors and larger tumors compared to uninfected mice and mice infected

with a CDT-deficient strain (He et al. 2019), demonstrating the tumorigenic properties of CDT.

A second example of a tumorigenic bacteria toxin is the *Bacteroides fragilis* toxin (BFT). The *bft* gene is more commonly found in the colonic mucosa of patients with tumors than those without tumors, suggesting a role for the toxin in tumorigenesis (Boleij et al. 2015). Mouse studies found that BFT induces tumors and is necessary for the tumor formation in those models (Jawara et al. 2019). Broadly speaking, BFT activates pro-inflammatory pathways and proliferation pathways that are generally associated with colorectal cancer (Cheng et al. 2020). One study found that BFT exposure alters gene expression and opens chromatin leading to changes in DNA methylation (Jawara et al. 2019). The sites of open chromatin and methylation are common in colorectal cancer, suggesting that BFT-induced epigenetic changes promote tumorigenesis.

A final example of a tumorigenic bacteria-produced toxin is colibactin, the toxin on which this thesis focuses. Colibactin is produced primarily by *Escherichia coli*, but can also be found in some *Klebsiella* species (Nougayrède et al. 2006). The toxin alkylates DNA, forming interstrand crosslinks that result in DNA double strand breaks upon removal of the lesion (Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). Repair of this DNA damage results in T>C mutations in an A/T-rich motif that can be found in patient tumor samples (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). In addition to mutations in driver genes, studies have found that colibactin exposure increases the tumor

burden in AOM/DSS treated mice (Cougnoux et al. 2014a, 2014b). Not only does colibactin increase tumor burden *in vivo*, but it is sufficient to transform healthy colon organoids by inducing constitutive Wnt signaling (Iftekhar et al. 2021).

1.2 Bacteria in the tumor microbiome

The presence of bacteria in tumors was first suggested over 100 years ago and since then many studies have aimed to characterize the tumor microbiome in a tissue-specific manner. One concern with claims that bacteria are present in tumors distant from the colon is the previously held idea that sites like the stomach or lungs are sterile. Thus, special care must be taken in these studies to ensure that contamination is not the source of bacteria. In fact, a recent paper was retracted due to flaws in the analysis that led to the conclusion that bacteria previously unassociated with humans were found in tumor samples (Poore et al. 2020; Abraham et al. 2023). Several works have classified the breast cancer microbiome by conducting 16S rRNA sequencing from breast tumor and adjacent healthy tissue. These studies identified similar bacterial composition in both the cancerous and non-cancerous tissues, but found higher abundance of bacteria in the tumor tissue (Camilla et al. 2016; Thompson et al. 2017). One study was able to identify differential microbial signatures depending on the type of breast cancer and even by patient outcome (Banerjee et al. 2018).

Perhaps the most comprehensive evaluation of bacteria in tumors comes from Nejman et al. (2020). The authors screened over 1,000 tumor samples and over 500 normal tissue samples with qPCR of the ribosomal 16S gene to detect

the presence of bacteria in each sample. They included several controls for contamination in their workflow and found varying frequencies of bacteria by tissue type (14.3% in melanoma compared to more than 60% in breast cancer). They also identified bacteria in tissues without exposure to the external environment that could explain the source of the bacteria, such as bone, ovarian, and brain tumors. To further support the validity of their finding, the authors used immunohistochemistry and RNA fluorescence in situ hybridization in tissue samples to identify bacterial cell walls and 16S rRNA, respectively. These stainings suggested that many bacteria are intracellular in both tumor and immune cells. The microbial profile of these tumor samples was then determined with sequencing to find that breast tumors in particular have both a higher abundance and diversity of bacteria. Additionally, each tissue type has a unique microbial profile that is in higher abundance in tumors compared to healthy tissue, as had previously been found in breast tissue.

1.3 Colibactin, a tumorigenic toxin

1.3.1 *Overview of colibactin*

Colibactin is a bacteria-produced genotoxin found predominantly in the B2 phylogroup of *E. coli* (Nougayrède et al. 2006; Wami et al. 2021). The toxin is encoded by a 54 kb genomic island, known as the *pks* island, containing 19 genes. These genes include non-ribosomal peptide synthetases and polyketide synthases, that together produce the colibactin toxin (Nougayrède et al. 2006). The genes comprising the *pks* island, and their known functions, are listed in Table 1.1.

Figure 1.1 shows a schematic of the colibactin production (Figure 1.1A) and export pipeline (Figure 1.1B). The toxin itself is composed of two cyclopropane rings, sometimes referred to as ‘warheads’, that directly bind DNA, forming interstrand crosslinks (Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). When colibactin was first discovered in 2006, it was hypothesized to play a role in tumorigenesis because of its ability to induce cell cycle arrest and DNA damage (Nougayrède et al. 2006).

Table 1.1 *pkis island genes*

Gene	type	function
clbA	PPTase	Colibactin biosynthesis phosphopantetheinyl transferase
clbR	Activator	Colibactin biosynthesis LuxR family transcriptional regulator
clbB	Hybrid NRPS/PKS	Colibactin hybrid NRPS/type I PKS
clbC	PKS	Colibactin PKS
clbD	Accessory	Colibactin biosynthesis dehydrogenase
clbE	NRPS	Colibactin biosynthesis aminomalonyl-acyl carrier protein
clbF	Accessory	Colibactin biosynthesis dehydrogenase
clbG	PKS	Colibactin biosynthesis acetyltransferase
clbH	NRPS	Colibactin NRPS
clbI	PKS	Colibactin PKS
clbJ	NRPS	Colibactin NRPS
clbK	Hybrid NRPS/PKS	Colibactin hybrid NRPS/type I PKS
clbL	Accessory	Colibactin biosynthesis amidase
clbM	Accessory	Precolibactin export MATE transporter
clbN	NRPS	Colibactin NRPS
clbO	PKS	Colibactin PKS
clbP	Peptidase	Precolibactin peptidase
clbQ	Accessory	Colibactin biosynthesis thioesterase
clbS	Accessory	Colibactin self-protection protein

Table adapted from Dougherty and Jobin 2021; Faïs et al. 2018

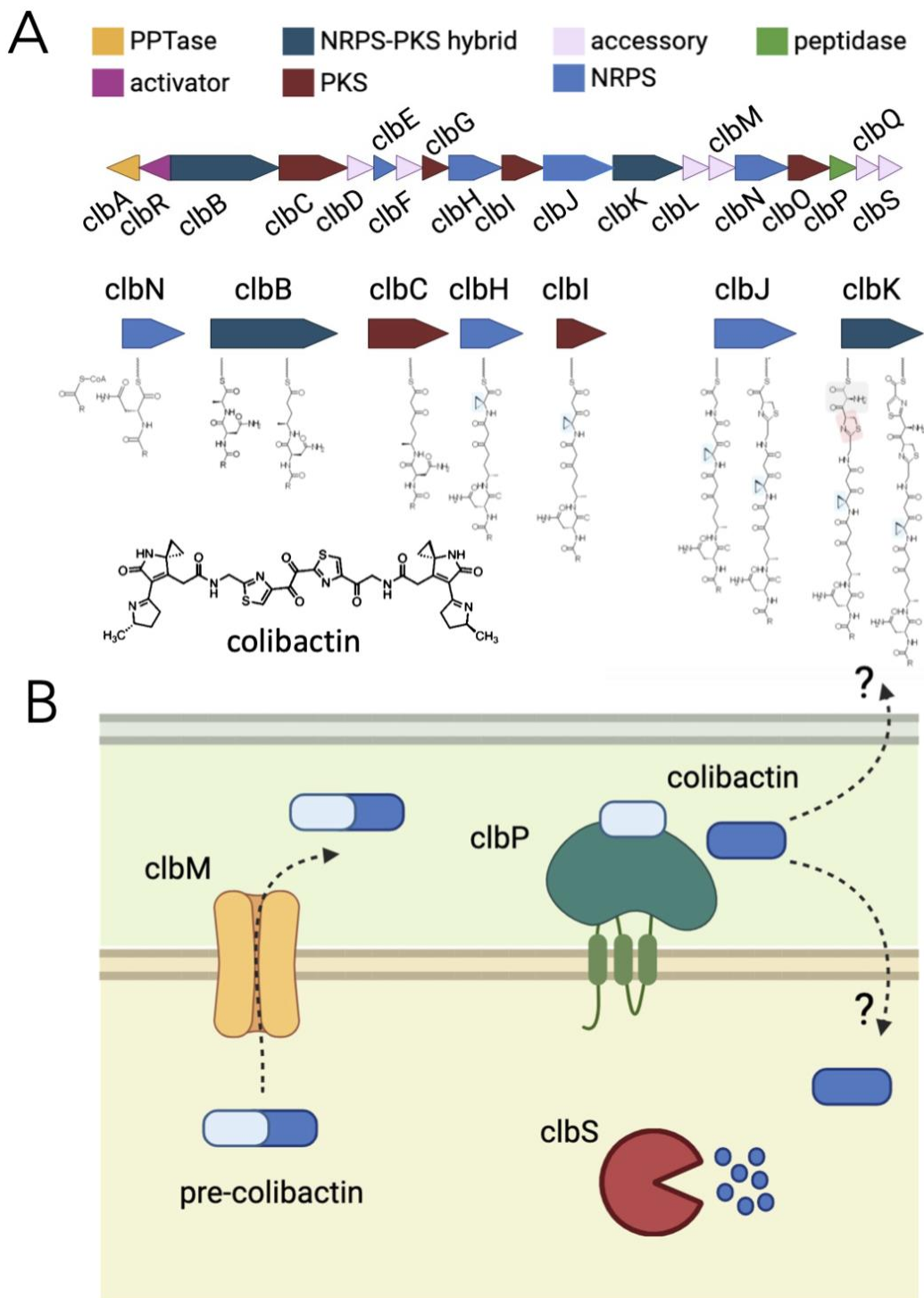


Figure 1.1 Colibactin is produced and exported by the pks island.

(A) Schematic of the 19 genes in the pks island along with their general functions (corresponding to Table 1.1). The bottom panel displays key steps in the production of colibactin (adapted from Fais et al. (2018)) and the chemical structure of colibactin (adapted from Dougherty and Jobin (2021)). (B) Schematic

of the final stages of colibactin production. Pre-colibactin is exported to the periplasm (in green) by the ClbM membrane channel. ClbP then makes the final cleavage to produce colibactin. Colibactin is then exported from the cell through a currently unknown mechanism. Some colibactin also re-enters the cytoplasm (in yellow), again through an unknown mechanism, where ClbS cleaves colibactin to protect the cell from DNA damage.

1.3.2 *Colibactin and its role in cancer*

Colibactin-producing bacteria have been associated with multiple human diseases ranging from inflammatory bowel disease to colorectal cancers (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). The toxin has been studied extensively in the context of colorectal cancer as the role of bacteria in promoting tumorigenesis has grown in the past decades. Initial studies of colibactin's tumorigenic properties focused on its ability to induce tumors in mouse models (Cougnoix et al. 2014a, 2014b) and its increased prevalence in colorectal cancer patients (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017).

Mutational and structural analyses provided mechanistic insight into how colibactin may promote tumorigenesis. Two key studies from 2020 identified a mutational signature induced by colibactin. The first of these studies cultured human colon organoids with *pks*⁺ bacteria isolated from a colorectal cancer (CRC) biopsy over a period of five months (Pleguezuelos-Manzano et al. 2020). A strain lacking *clbQ* that is unable to produce colibactin was used as a control. Whole-genome sequencing identified T>N mutations in ATA, ATT, and TTT sequences in colibactin-exposed organoids, but not in the control organoids. This was assigned to the COSMIC database mutational signature SBS88, along with single T deletions assigned ID18. These two mutational signatures were then

found to be overrepresented in CRC-derived metastases compared to all other cancers. In addition to identifying mutations associated with colibactin exposure in patient samples, the driver gene *APC* had more than 5% of its driver mutations associated with the A/T rich colibactin motif.

The second paper used an alternative approach to identify the colibactin-induced mutational signature. The authors exposed a colorectal cancer cell line to pks⁺ or pks⁻ *E. coli* for three hours before collecting cells for in-suspension breaks labeling in situ and sequencing (sBLISS) (Dziubańska-Kusibab et al. 2020). sBLISS captures DNA sequences with double strand breaks (DSBs) for sequencing. Motif discovery tools identified an enriched A/T hexanucleotide motif where the DSBs occurred in colibactin-exposed cells. Structural analysis of all possible hexanucleotides found that the A/T hexanucleotides (AAAATT/AATTTT and AAATTT) had the narrowest minor grooves, strong negative electrostatic potential, and extreme rigidity. Modeling of colibactin in these minor grooves showed stable binding with the cyclopropane rings near adenines, which is logical as colibactin has been shown to alkylate adenines (Wilson et al. 2019). In addition to the structural evidence that colibactin binds AAAATT/AATTTT and AAATTT motifs, analysis of tumor datasets found these motifs enriched compared to random A/T motifs. Again, mutations in the driver gene *APC* were found. Taken together, these two studies provide key evidence for the role colibactin plays in tumorigenesis and the genomic footprint its toxicity leaves in cancers. These studies also inform the potential DNA repair pathways that may respond to colibactin-induced DNA damage.

1.4 Bacteria-bacteria competition

1.4.1 *Overview of microbial competition*

Competition between bacteria can arise when nutrients or space are limited. This competition can be within a species, or across species in a larger community. Bacteria take two general approaches to increasing their survival in competitive environments (Hibbing et al. 2010; Cornforth and Foster 2013). The first is through nutrient uptake, otherwise termed exploitation, which increases nutrient access for the uptaking cell, but reduces environmental levels of a nutrient for competing cells. The second is through warfare, otherwise termed interference, which includes the production of toxins or other damaging agents to kill competing bacteria.

1.4.2 *Nutrient uptake as an approach to competition*

Bacteria can use exploitation in a competitive environment by uptaking certain nutrients or by producing specialized metabolites that harvest these nutrients. *Bacteroides thetaiotaomicron* has one of the largest gene sets for acquiring and metabolizing carbohydrates (Keeney and Finlay 2011). These include surface proteins to bind and import carbohydrates and hundreds of enzymes, including glycoside hydrolases and polysaccharide lyases, to metabolize carbohydrates (Bäckhed et al. 2005). These nutrients are essential for *B. thetaiotaomicron* to stably colonize the host gut, as it relies heavily on glycoproteins on its cell membrane to prevent wash-out from the gut (Bäckhed et al. 2005). The pathogenic *E. coli* O157:H7 strain contains genes to catabolize

maltose, which is essential for its colonization of the gut when competing with commensal *E. coli* (Jones et al. 2008; Keeney and Finlay 2011). *E. coli*, in general, need glycogen for host gut colonization, and thus have mechanisms of storing internal glycogen to help survive outside of the host and persist in the colonized host (Jones et al. 2008).

Specialized metabolites can be produced to acquire certain nutrients. One of the most common of these metabolites is siderophores, which help with iron-scavenging (Hibbing et al. 2010). Siderophores are costly to produce, but are essential in low-iron environments and thus are regulated by iron. Siderophores have a high (but varying) affinity for iron (Figure 1.2, examples 1 and 3), allowing them to easily bind iron molecules and chelate them (Keeney and Finlay 2011; Wandersman and Delepelaire 2004; Joshi et al. 2006). The siderophore must then interact with the cell membrane to transport the iron into the cell. Because siderophores must leave the cell and then return with iron, many bacteria can interact with heterologous siderophores (Joshi et al. 2006; Keeney and Finlay 2011). This practice introduces additional strategies to compete for nutrients (Figure 1.2). Some bacteria, upon interacting with a heterologous siderophore, will upregulate their own siderophore production (Figure 1.2, example 4) (Keeney and Finlay 2011). Others solely rely on heterologous siderophores to assist in their iron uptake (Figure 1.2, example 2), but this depends on the relatedness of competing species, as there is a wide variety of siderophores produced (West and Buckling 2003; Harrison et al. 2008). While this last strategy may appear risky

due to the reliance on other bacteria, it saves the cell from the cost of producing siderophores.

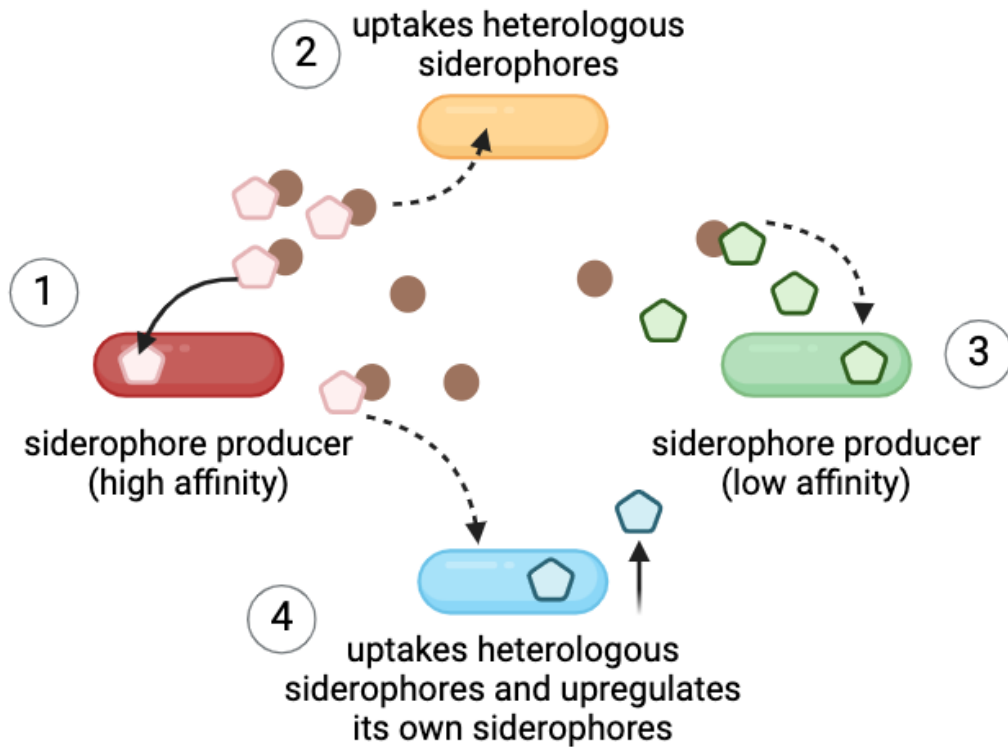


Figure 1.2 Bacteria take varying approaches to iron uptake.

Schematic of different approaches to iron uptake via siderophores, a specialized metabolite. Bacteria 1 produces high affinity siderophores (pink). Bacteria 2 can uptake heterologous siderophores, such as those produced by bacteria 1. Bacteria 3 is a low affinity siderophore producer (green), which can scavenge iron, but at a lower efficacy. Bacteria 4 can uptake heterologous siderophores (pink), which then upregulates expression of its own siderophores (blue). Iron is represented by brown circles.

1.4.3 *Bacteria warfare*

Interference in bacterial competition involves damaging competitor bacteria cells. Bacteria produce and secrete damaging agents, including antibiotics. These approaches can be contact dependent, such as the type 6 secretion system (T6SS), or function over longer distances, like bacteriocins.

Pseudomonas aeruginosa harbors both systems: a T6SS and tailocins (Booth et al.

2023). The T6SS delivers effector proteins into recipient cells, which can be both host and bacteria cells. T6SSs aid in infecting host cells, while bacterial cells are killed by T6SSs. The toxins delivered by these systems vary depending on the producing cell species. One T6SS effector in *P. aeruginosa* is a metalloprotease that enters the target cell periplasm, causing morphological changes to the cell membrane that lead to toxicity (Wood et al. 2019). Tailocins are bacteriocins similar in structure to the tail of a bacteriophage and induce membrane depolarization of targeted cells, causing cell death. The production of tailocins leads to the lysis of the producing cell, but only a few molecules are needed to kill competing cells (Booth et al. 2023; Carim et al. 2021). Tailocin sensitivity is mediated by the O-antigen composition of cell membranes (Carim et al. 2021). Interestingly, contact-dependent killing mechanisms are most effective at intermediate attack frequencies, while long-range killing approaches are most effective at high frequency attacks (Booth et al. 2023).

1.4.4 *Colibactin as mechanism of interference competition*

The colibactin toxin was discovered in 2006 for its ability to induce cell cycle arrest and DNA damage (Nougayrède et al. 2006). While initial work focused on the potential role for colibactin in tumorigenesis, some work hypothesized that the toxin may play a role in bacterial competition. The genomic island has a module that produces thiazole rings, commonly found in natural products like bleomycin (Nougayrède et al. 2006). The colibactin pathogenicity island encodes a cyclopropane hydrolase, ClbS, that cleaves the cyclopropane

warheads on the toxin (Nougayrède et al. 2006; Bossuet-Greif et al. 2016; Tripathi et al. 2017). This protein serves as a self-protector for colibactin-producing strains, which suggests that colibactin would otherwise induce auto-toxicity. In fact, cells lacking the *clbS* gene along with the pathway for nucleotide excision repair showed self-toxicity (Tripathi et al. 2017; Bossuet-Greif et al. 2016). Further work connected colibactin to siderophore production, suggesting that the toxin is upregulated in iron-limited environments. *E. coli* have a PPTase in the core genome that is required to make siderophores, but the colibactin pathogenicity island also has a PPTase, ClbA (Martin et al. 2013). In a pathogenic *E. coli* strain, deletion of both PPTases resulted in a loss of virulence, whereas expression of either one of the PPTases was sufficient for infection in a mouse model (Martin et al. 2013). These experiments suggest a close relationship between an exploitation approach and possible interference approach to competition where redundancies in the pathways may protect cells from losing all competitive approaches entirely.

1.4.5 *Colibactin toxicity to neighboring bacteria*

As previously mentioned, colibactin may serve as a toxin to mediate interference competition, but this role was not investigated for nearly a decade after its initial discovery. The first study to investigate this question evaluated growth inhibition of clinical strains by colibactin-producing strains. These clinical strains covered dozens of species. The authors found that colibactin causes growth inhibition of *Staphylococcus aureus*, including multi-drug resistant strains (Faïs et

al. 2016). This concept was not further expanded upon until several papers were published in 2022.

Silpe et al. (2022) evaluated colibactin toxicity to single strains and communities. The authors found that colibactin induces prophage excision leading to cell lysis across several bacteria strains and different prophages. This observation was also made for a community of gut strains cultured *ex vivo* with colibactin-producing bacteria. In addition to identifying a mechanism of colibactin toxicity, this study identified resistant proteins in non-colibactin producing bacteria. Within *E. coli*, seven strains were identified containing the *clbS* gene and a truncated copy of the gene immediately upstream, *clbQ*. There were also transposase-associated genes flanking the region, suggesting horizontal gene transfer. More than 200 ClbS-like proteins were also identified in distantly related species with 20-80% homology, ten of which were tested and provided colibactin resistance. Wong et al. (2022) showed colibactin toxicity to *S. aureus*, as others showed, but demonstrated this in prophage-cured strains, contradicting the conclusion of Silpe et al. (2022) that prophage excision is the mechanism through which colibactin induces toxicity. An additional group investigated *Vibrio cholerae* colonization of the mouse gut and found that *E. coli* producing colibactin inhibited gut colonization and induced the SOS response (DNA damage response) (Chen et al. 2022). Lastly, it was also found that the prodrug motif cleaved during the final step of colibactin production is *N*-myristoyl-D-Asn, which has shown modest growth inhibition in *Bacillus subtilis*. While *N*-myristoyl-D-Asn does not cause growth inhibition in *S. aureus*, the authors did

find that it permeabilizes its membrane (Wong et al. 2022). The mechanisms for colibactin cell entry has yet to be elucidated, but this finding suggests that the toxin may simply diffuse into target cells.

1.5 DNA damage repair pathways in humans and *E. coli*

1.5.1 Overview of DNA damage repair

There are many pathways involved in the DNA damage response and these are highly conserved, for the most part, across species at the biochemical level. The number of proteins involved can vary and increases with species complexity. Initial understanding of DNA damage repair comes from studies in *E. coli* and provided a model that could be built upon in more complex organisms, like humans. Through different stages of the cell cycle, the base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), non-homologous end joining (NHEJ) (predominantly in eukaryotes), and homologous recombination (HR) pathways are active (Chatterjee and Walker 2017). Interstrand cross link (ICL) repair and translesion synthesis (TLS) are activated in response to DNA lesions. Each type of DNA damage, and each damaging agent within a class of agents, activates different repair pathways (Table 1.2). DNA polymerases have error rates ranging from 10^{-1} to 10^{-7} (Chatterjee and Walker 2017), supporting the concept that certain pathways are more error-prone than others, as each pathway may use different polymerases. Many cancers have malfunctioning repair pathways, which contributes to further mutagenesis and genomic instability.

Table 1.2 Repair pathways activated by types of DNA damage

Damaging agent class	DNA repair pathways
Irradiation	HR
Ultra-violet	NER, ICL repair, TLS, HR
Alkylation	BER, ICL repair
Aromatic amines	NER
Polycyclic aromatic hydrocarbons	TLS, NER, BER

1.5.2 *Base excision repair*

BER is a major repair pathway that functions by recognizing damage with DNA glycosylases, where there are unique DNA glycosylases for each base (Chen et al. 2024; Matsumoto et al. 1999; Krokan et al. 2000). In humans, glycosylases for oxidized bases may have some redundancies to protect cells from extensive DNA damage in the case of functional loss of one protein (Hazra et al. 2007). The DNA glycosylase cleaves the damaged base from the sugar backbone, leaving an apurinic/apyrimidinic (AP) site. AP endonuclease recognizes the AP site and generates a single strand break. DNA polymerase then fills the gap and DNA ligase repairs the nick in the DNA backbone to complete the repair process.

1.5.3 *Nucleotide excision repair*

NER has two main pathways: global genome repair (GGR), which can repair any damage in the genome causing DNA helix distortions, and transcription-coupled repair (TCR), which repairs damage blocking the progression of transcription (Fousteri and Mullenders 2008; Vermeulen and Fousteri 2013; Wood 1997; Marteijn et al. 2014). For TCR-NER, RNA polymerase, in *E. coli* (Selby and Sancar 1990), and RNA polymerase II (Selby et

al. 1997), in humans, backtracks after encountering a lesion to allow for repair. After each pathway recognizes the DNA damage, endonucleases are recruited to cleave DNA on both sides of the lesion to excise the damage. This excised region is 12-13 nucleotides in prokaryotes (Sancar and Rupp 1983) and 26-29 nucleotides in eukaryotes (Huang et al. 1992). DNA polymerase then fills the gap using the undamaged strand as a template and DNA ligase seals the nick to finish the repair process. In the case of TCR-NER, RNA polymerase can continue replication after this step. Removing the lesions during NER can activate the ATR-mediated DNA damage response that modifies the H2A and H2A.X histones in humans. The entire pathway consists of three proteins in *E. coli* and 15-16 in humans (Hu et al. 2017).

1.5.4 *Mismatch repair*

MMR typically functions during DNA replication to repair any errors made by DNA polymerase that generate mismatching nucleotides or insertion/deletion loops. This is a highly conserved process across prokaryotes and eukaryotes, with deficiencies in the pathway associated with hypermutator states. MMR increases DNA replication accuracy 100-1000 fold, so loss of this pathway significantly increases the mutational rate (Buermeyer et al. 1999; Li 2008; Bellacosa 2001; Modrich 1997; Schofield and Hsieh 2003). MMR in *E. coli* involves “mutator” genes, termed mutator genes as their loss increases the mutational rate, exonucleases, single-stranded binding protein, DNA polymerase, and DNA ligase (Li 2008). Many of the components of human MMR are

homologs of these *E. coli* proteins. MMR relies on strand signals from within the DNA helix that differentiate daughter and parental strands. In *E. coli* this relies on the lack of hemi-methylated dGATC sites on newly synthesized daughter strands that guides nicking. The method of strand differentiation in humans is still under investigation, but appears to overlap with the process in *E. coli* by involving the identification of DNA nicks (Buermeier et al. 1999; Li 2008; Putnam 2021). Once the mismatch is identified, an exonuclease excises the mismatch. DNA polymerase then fills the gap and DNA ligase seals the final nick to complete the repair process.

1.5.5 *Non-homologous end joining*

Non-homologous end joining is responsible for repairing up to 80% of DSBs in human cells (Karanam et al. 2012). *E. coli* predominantly use HR to repair DSBs, but some strains possess an alternative end joining repair (A-EJ) pathway that does not depend on the key NHEJ proteins in humans (Chayot et al. 2010). Upon DSBs that require NHEJ in humans, both of the broken ends of the DNA are captured by Ku proteins, which recruit additional enzymes (Barnes 2001; Weterings and Chen 2008; Chang et al. 2017; Stinson and Loparo 2021; Karanam et al. 2012). These enzymes can be nucleases to trim the ends or polymerases to fill in missing nucleotides to ensure there is enough homology between the ends. A DNA ligase can then ligate the two DNA ends to repair the DNA. Because nucleotides may need to be excised to get sufficient homology for ligation, NHEJ is more likely to result in mutations despite tight regulation. A-EJ

in *E. coli* involves resection by a nuclease/helicase complex (*recBCD*) rather than end protection by Ku proteins that recruit strand processing enzymes (Chayot et al. 2010). After strand resection, microhomology (1-9 nucleotides) is used to ligate the strands with a ligase that differs from the ligase in human NHEJ. This repair is mutagenic because of the strand resection and microhomology use (Chayot et al. 2010).

1.5.6 *Homologous recombination*

HR is involved in repairing DSBs and ICLs and can help recover stalled replication forks (Sun et al. 2020; Jasin and Rothstein 2013; Li and Heyer 2008; Vasquez et al. 2001; Wright et al. 2018). HR is the main pathway for repairing DSBs and is conserved across prokaryotes and eukaryotes. When a DSB is introduced, the DNA strand ends must be recognized and processed to a single stranded tail. This processing can include trimming by exonucleases. Rad51, in humans, or RecA, in *E. coli*, then forms a filament with the single strand DNA (Kuzminov 2011) and does strand invasion to form a D loop. This strand invasion provides a template for DNA repair synthesis. The damaged strands can then be ligated, sometimes with crossing over to exchange material between sister chromatids and sometimes without.

1.5.7 *Other repair pathways*

There are other pathways for DNA repair that may be made up of several pathways previously described. These pathways are studied mostly for their role

in cancer and include TLS, Fanconi Anemia repair, and ICL repair. TLS is a form of DNA damage tolerance. Cells can use TLS to bypass DNA damage lesions with error-prone DNA polymerases, which are conserved across kingdoms, leading to mutation accumulation (Waters et al. 2009). Fanconi Anemia (FA) repair is a pathway first characterized by genes associated with hypersensitivity to crosslinking agents and a predisposition for cancer. This repair pathway exists in humans, but not *E. coli*. Some of these proteins interact with HR proteins and single strand annealing proteins (a pathway that anneals complementary strands after a DSB) (Nakanishi et al. 2005). A key step in FA repair is the ubiquitination to activate the pathway and lead to nucleolytic processing of the lesion to unhook DNA. From there, other pathways like TLS and HR are activated to complete the DNA repair (Moldovan and D'Andrea 2009). ICL repair is a coordinated response involving several DNA repair pathways. NER proteins identify the lesion and introduce nicks to the DNA on either side of the lesion to excise it (Dronkert and Kanaar 2001; McHugh et al. 2001; Muniandy et al. 2010; Deans and West 2011). Here, the Fanconi Anemia proteins may be recruited to the lesion and further activate HR. In *E. coli* lacking the FA pathway, the repair continues straight to HR involvement after NER proteins excise the lesion (Wood 2010). TLS may also play a role in bypassing the DNA lesions. If a lesion is encountered during transcription, RNA polymerase may stall and activate TC-NER.

1.6 Mutational biases

1.6.1 *Mutation bias in human cells from DNA repair*

Mutational bias from certain repair pathways is important to understand to determine which repair pathways respond to certain types of DNA damage. Across eukaryotes, selection appears to favor G/C>T/A transitions, which may be due to stop codons being A/T rich (Lynch 2010). In terms of strand symmetry, mutations in human cells favor A>G mutations over T>C, especially in transcribed regions, in the case of single base substitutions, which may be caused by TC-NER (Green et al. 2003). One group analyzed a dataset of single nucleotide polymorphisms (SNPs) occurring in “neutral” regions of the genome, meaning the SNPs were outside of exons, regulatory regions, splice sites, untranslated regions, and CpG islands. From this analysis, the authors found that transitions (C:G>T:A or A:T>G:C) represented over half of the mutations (Nakken et al. 2010). One of the most frequent mutations occurs in CpG islands, where a methylated cytosine spontaneously undergoes deamination, resulting in a thymine and generating a G:T mismatch (Poulos et al. 2017). These mutations are seen in many cancer types, but can also be seen in non-cancer tissues, as they increase with age. Exogenous agents, like UV and cigarette smoke, can also induce deamination of methylated cytosines.

Most of our insight into mutational biases comes from studies in cancer cells deficient in specific pathways. Bladder cancers lacking NER show a higher level of mutations across all nucleotide changes (Guo et al. 2013; Kim et al. 2016). Cancer cells with somatic hypermutations, which can be driven by UV,

alcohol, and other mutagens, activate an error-prone MMR pathway that uses a TLS polymerase, leading to frequent A>C or A>G mutations (Supek and Lehner 2017). A mapping of mutational signatures in cancers revealed C>A mutations on the transcribed strand were associated with TC-NER removing DNA adducts of polycyclic hydrocarbons from tobacco smoke (Alexandrov et al. 2013). This study also identified C>T mutations on the untranscribed strand that are consistent with damage from UV that is repaired with TC-NER. In MMR deficient cancers, a C>T signature in NpCpG mutations was found (Alexandrov et al. 2013).

1.6.2 *Mutation bias in E. coli*

Many studies have evaluated spontaneous mutations in *E. coli*, partially due to the short generation time, which allows for quicker accumulation of mutations. Good et al. (2017) passaged *E. coli* over 30 years and for more than 60,000 generations to understand how mutations arise and stabilize over time. Of the 12 lines grown over time, several developed mutations in DNA repair pathways, making the lineages hypermutator populations. In non-mutator lines, mutations were distributed non-randomly. Further analysis of this dataset revealed that hypermutator strains lost either NER or MMR (Maddamsetti and Grant 2020). Several of the hypermutator strains that were MMR-deficient favored A:T>G:C and G:C>A:T mutations. In terms of distribution of mutations, MMR-deficient hypermutator strains displayed a wave pattern in mutations, where the least mutations occurred at the origin of replication and the most occurred

between the origin and terminus (Maddamsetti and Grant 2020). Interestingly, mutT mutants showed even distribution of mutations across the genome.

Additional studies have been conducted over shorter time-spans. One study over 6,000 generations in a wild-type or hypermutator (MMR-deficient) *E. coli* strain found that under minimal selective pressure, *E. coli* accrue 1×10^{-3} mutations per genome per generation (Lee et al. 2012). This work found that wild-type strains accumulated mutations evenly distributed across the genome, while hypermutators accumulated mutations that were not evenly distributed. The wild-type strain showed A:T>C:G and G:C>T:A mutations making up more than half of the observed mutations, while these transitions represented 98% of the mutations in the hypermutator strain. The local sequence context influenced these mutations, with **A**C and G**T** sites (mutated base in bold) accounting for 79% of the A:T>C:G mutations. Overall, mutation accumulation experiments that do not put selective pressure on the *E. coli* favor transition mutations (A:T>C:G and G:C>T:A) (Lee et al. 2012; Maddamsetti and Grant 2020; Good et al. 2017).

1.6.3 *Damage-induced mutational signatures*

One of the most studied mutagens is ultraviolet (UV) exposure. In both humans and *E. coli* UV induces a majority of C>T mutations at CC or CT sites (Brash 2015; Pfeifer et al. 2005; Todd and Glickman 1982). A study investigating environmental mutagens found that aristolochic acid I, dibenzo[*a,l*]pyrene diol-epoxide, and dibenzo[*a,l*]pyrene, which all form adenine adducts, induced A:T>T:A mutations in human cells (Kucab et al. 2019). In *E. coli*, dibenzo[*a,l*]pyrene diol-epoxide induces both G:C>T:A and A:T>T:A

transversions (Bernelot-Moens et al. 1990), suggesting that mutational differences exist between *E. coli* and humans. Cisplatin, an alkylating agent and chemotherapy, induces A:T>T:A and G:C>T:A transversions in *E. coli* (Burnouf et al. 1987). Human cancer cell lines exposed to cisplatin predominantly showed C>T mutations in **C**CC or C**C**T sequences (mutated base in bold) (Boot et al. 2018), showing some overlap with mutations observed in *E. coli*. Given that both human and *E. coli* both favor A:T>G:C or C:G>T:A transition mutations without selective pressure, it is interesting that DNA damage results in differing mutations. This could be due to activation of different DNA repair pathways between the two species, or to different biases in errors made during DNA repair.

1.7 Scope of this thesis

While much progress has been made to understand the underlying mechanisms of colibactin toxicity in the past two decades, there are still several outstanding questions. The mechanism through which colibactin is secreted from producing cells has evaded researchers, along with how it enters target cells. Recent work found evidence that the toxin may enter via diffusion (Wong et al. 2022), but past works also found colibactin present in outer membrane vesicles (Cañas et al. 2016; Fábrega et al. 2017). If colibactin is secreted from cells in outer membrane vesicles, a genetic screen should detect resistance in cells lacking expression of genes involved in endocytosis. Colibactin toxicity has been demonstrated in both mammalian and bacteria cells, so there is likely a similar

path of cellular entry, but there is still the possibility that it differs between prokaryotes and eukaryotes.

Some work has investigated the DNA damage response in mammalian cells following colibactin exposure, but these works have not systematically evaluated all DNA repair pathways. Few studies have investigated the bacteria cell response to colibactin, other than to confirm that there is DNA damage (Chen et al. 2022; Wong et al. 2022). Again, while these responses are likely similarly across kingdoms due to the conservation of DNA repair pathways, there still remains a possibility that these responses diverge. Additionally, a mutational signature is induced in human cells after colibactin exposure, but we do not know if bacteria exhibit the same signature, or if they experience mutations after colibactin-induced damage. Much of the work in mammalian cells needs to be repeated in bacteria models to better understand how universal colibactin toxicity is. This thesis aims to answer these outstanding questions and shed more light on underlying mechanisms of colibactin toxicity in both mammalian cells and bacteria.

Chapter 2: Systematic mapping of the DNA damage response to colibactin exposure in colon cancer

2.1 Abstract

Colibactin is a toxin produced primarily by the B2 phylogroup of *Escherichia coli* that crosslinks DNA in host cells. This results in double strand breaks that are repaired erroneously, introducing mutations in host colon epithelial cells. Some pathways involved in repairing colibactin-induced DNA damage have been identified or proposed. However, systematically identifying the DNA repair pathways would provide more insight into the mechanisms of colibactin toxicity in host cells and allow for comparisons to other alkylating agents. We established that colibactin induces increasing DNA damage at higher bacteria loads in a colon cancer cell line. We then validated that these bacteria loads induce transient cell cycle arrest, allowing for future cell proliferation. Using these bacteria load titrations, we employed a genome-wide CRISPR screen to identify the DNA repair pathways activated by colibactin under different levels of colibactin exposure. We identified roles for nearly all DNA repair pathways in responding to colibactin-induced damage, with especially strong involvement of the Fanconi anemia/interstrand crosslink repair and fork quality control pathways. This work reveals cellular responses to colibactin toxicity and lays the groundwork for future co-cultured genetic screens in cancer cells.

2.2 Introduction

The presence of microbes in the tumor environment has been established for decades. However, the composition of these species and the role bacteria play in the tumor microenvironment are still under investigation. Several tumorigenic bacteria have been identified, including bacteria that produce colibactin. Because of its implication in colon cancer, colibactin has been extensively studied in the context of host cell DNA damage. While many aspects of colibactin toxicity have been discovered, its mechanism of cellular entry and the DNA damage response to its toxicity have yet to be identified. Addressing this knowledge gap could identify potential targets for inhibitors to protect host colon cells from colibactin and suggest similarities between the cellular response to colibactin and other well-studied DNA damaging agents. We systematically evaluated colibactin toxicity in colon cancer cells. We found a key role for the Fanconi anemia/interstrand crosslink repair and fork quality control pathways in repairing colibactin-induced DNA damage. We were unable to identify a route of cellular entry, but this may support one hypothesis that colibactin diffuses into target cells.

Colibactin is a toxin produced primarily by the B2 phylogroup of *E. coli* that binds and crosslinks DNA, resulting in double strand breaks (Nougayrède et al. 2006; Wami et al. 2021; Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). The toxin is encoded by the pks pathogenicity island and bacteria containing this island are found more prevalently in colorectal cancer (CRC) patients than

healthy individuals (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). Colibactin was originally hypothesized to be a potentially tumorigenic agent based on its ability to induce cell cycle arrest and DNA damage (Nougayrède et al. 2006). The pks pathogenicity island is 54 kb and contains 19 genes that together produce the colibactin toxin (Nougayrède et al. 2006). The toxin itself has two cyclopropane rings on each end that directly bind and crosslink DNA (Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). Cells activate multiple repair pathways to resolve colibactin crosslinks, including non-homologous end joining (Cuevas-Ramos et al. 2010), homologous recombination (Dougherty et al. 2023), and Fanconi anemia repair pathways (Bossuet-Greif et al. 2018; Dougherty et al. 2023). This list may not be exhaustive, as these studies each primarily targeted a single DNA repair pathway. Following removal of the crosslink and DNA repair, cells exhibit a mutational signature specific to colibactin (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). These mutations occur throughout the genome and can be found in driver genes, such as APC (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020; Rosendahl Huber et al. 2024).

While much progress has been made to understand the cellular response to colibactin-induced damage, an approach interrogating all DNA repair pathways, including genes that may play roles in multiple pathways, has yet to be conducted. To address this gap, we conducted a whole-genome CRISPR screen in a single

CRC cell line to maintain a consistent genetic background. We found that Fanconi anemia/interstrand crosslink repair and fork quality control play central roles in responding to colibactin toxicity. Our screen also lays the groundwork for future genetic screens that involve co-culturing host cells with extracellular bacteria to determine mechanisms of toxicity.

2.3 Results

2.3.1 Colibactin induces dose-dependent DNA damage

We first established that colibactin-producing bacteria induce DNA damage in a CRC cell line. We co-cultured the HCT116 CRC cell line with wild-type Nissle 1917 (EcN) or a *clbN* deletion that does not produce colibactin (EcN $\Delta clbN$). These co-cultures were conducted over a range of bacteria loads to validate that colibactin induces DNA damage in our co-cultures and to determine if this is a dose-dependent effect. After four hours, bacteria were washed off and cells were fixed for immunofluorescent staining with an anti- γ -H2A.X antibody (Figure 2.1A). Cells phosphorylate the H2A.X histone upon recognition of DNA double strand breaks (Rogakou et al. 1998), so we were able to use this as a marker of DNA damage. We observed generally brighter γ -H2A.X signal (red) and more foci in the nuclei of EcN treated cells compared to EcN $\Delta clbN$ treated cells (Figure 2.1B). This signal was heterogenous between cells within replicates for both conditions.

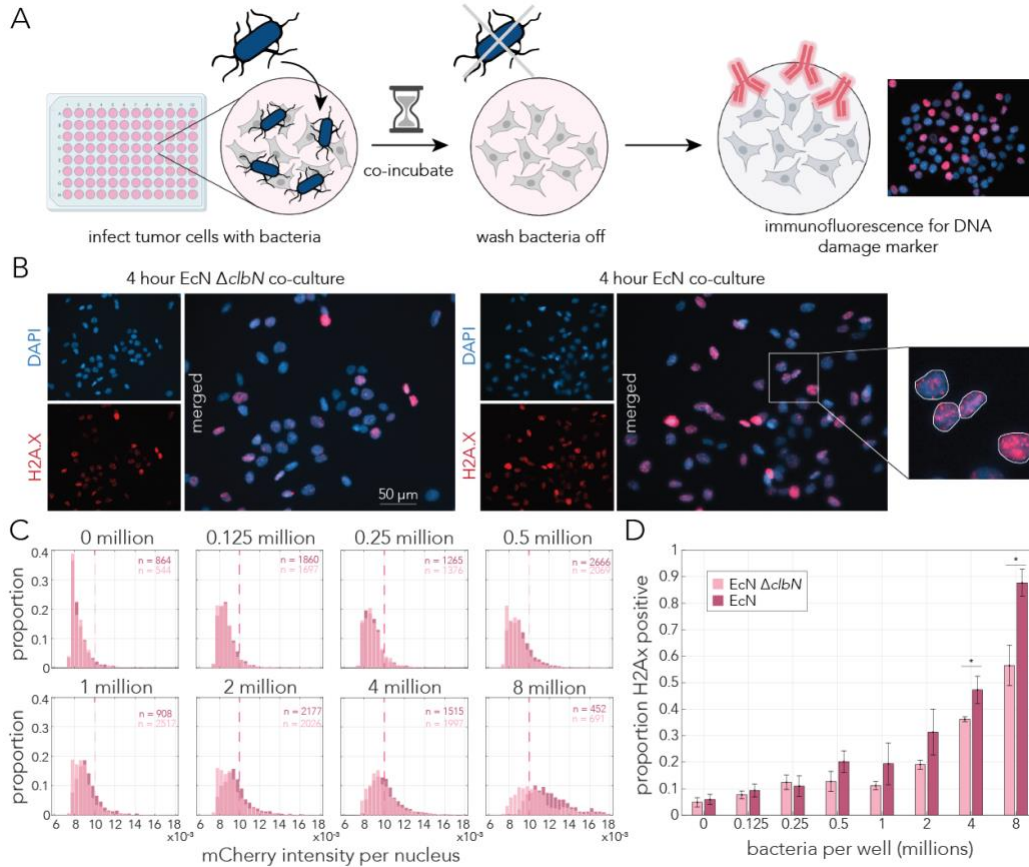


Figure 2.1 Colibactin induces DNA damage in a dose-dependent manner.

(A) Schematic of colibactin and HCT116 co-culture and immunofluorescent staining. Cells are co-cultured and then bacteria are washed off and the tumor cells are stained with an antibody that indicates DNA damage. (B) Representative images of EcN and EcN $\Delta clbN$ exposed tumor cells. The blue channel is a DAPI nuclear stain and the red is γ -H2A.X. The righthand inset shows example nucleus segmentation for signal intensity analysis. (C) Histograms of mean AF647 intensity (γ -H2A.X) per nucleus per bacteria load (reported in millions per well). EcN (colibactin) is shown in dark pink and EcN $\Delta clbN$ (no colibactin) is shown in light pink. The number of nuclei analyzed per bacteria load are written in the same colors on each plot. Replicates were pooled for each bacteria load to plot in the histograms. Vertical dashed lines represent the cutoff of 0.001 used in panel D for γ -H2A.X positive cells. (D) Mean proportion of γ -H2A.X positive cells for EcN and EcN $\Delta clbN$ exposed tumor cells across bacteria loads. Each condition was conducted in triplicate (with the exception of EcN $\Delta clbN$ 0.25 million where one replicate had high background signal and was unable to be quantified) and error bars represent standard deviation. * $p < 0.05$

We quantified the amount of DNA damage by segmenting nuclei using the DAPI nuclear stain. Within each nucleus, we quantified the mean red (γ -H2A.X) signal intensity. Figure 2.1C shows histograms of the signal intensity per nucleus for both EcN and EcN $\Delta clbN$ exposed tumor cells. Overall, as the number of bacteria increased, the amount of DNA damage also increased, regardless of whether the bacteria were producing colibactin. However, the peak of the distribution for colibactin-exposed tumor cells appeared to be shifted to the right (more DNA damage) compared to the control tumor cells. To confirm this, we set a cutoff signal intensity to determine if a nucleus was γ -H2A.X-positive. Figure 2.1D shows the proportion of DNA damage positive cells for both EcN and EcN $\Delta clbN$ treated tumor cells over each bacteria load. While the amount of damage clearly increases for both bacteria conditions, there is a significant difference when tumor cells were co-cultured with 4×10^6 and 6×10^6 bacteria. These results agree with the bacteria load previously used for EcN in co-cultures (Hallam et al. 2023; Nougayrède et al. 2021). Taken together, these results validate past observations of colibactin-induced DNA damage in tumor cells and show that the presence of bacteria alone can induce DNA damage, likely from stress.

2.3.2 Colibactin induces cell cycle arrest in a dose-dependent manner

Due to the elevated DNA damage solely from the presence of bacteria, we next wanted to evaluate cell cycle arrest after colibactin exposure to determine if this phenotype is exclusively due to colibactin. Figure 2.2A outlines the general workflow of these experiments. We again co-cultured tumor cells with either EcN

or EcN $\Delta clbN$ with different bacteria loads and a no bacteria control. After a four-hour co-culture, the bacteria were washed off and the replenished media was supplemented with antibiotics to kill any remaining bacteria and prevent further growth. Tumor cells expressed an mCherry histone tag that caused a red fluorescent nucleus. We imaged the tumor cells with timelapse microscopy over a period of six days to monitor their growth. The nuclear fluorescence allowed us to count the number of cells per well. Figure 2.2B shows the masked nuclei in representative images for an EcN, EcN $\Delta clbN$, and no bacteria well every day for six days. While all three conditions show some degree of proliferation, the no bacteria and EcN $\Delta clbN$ conditions show approximately the same amount of growth. The EcN exposed condition had fewer nuclei, even six days after colibactin exposure.

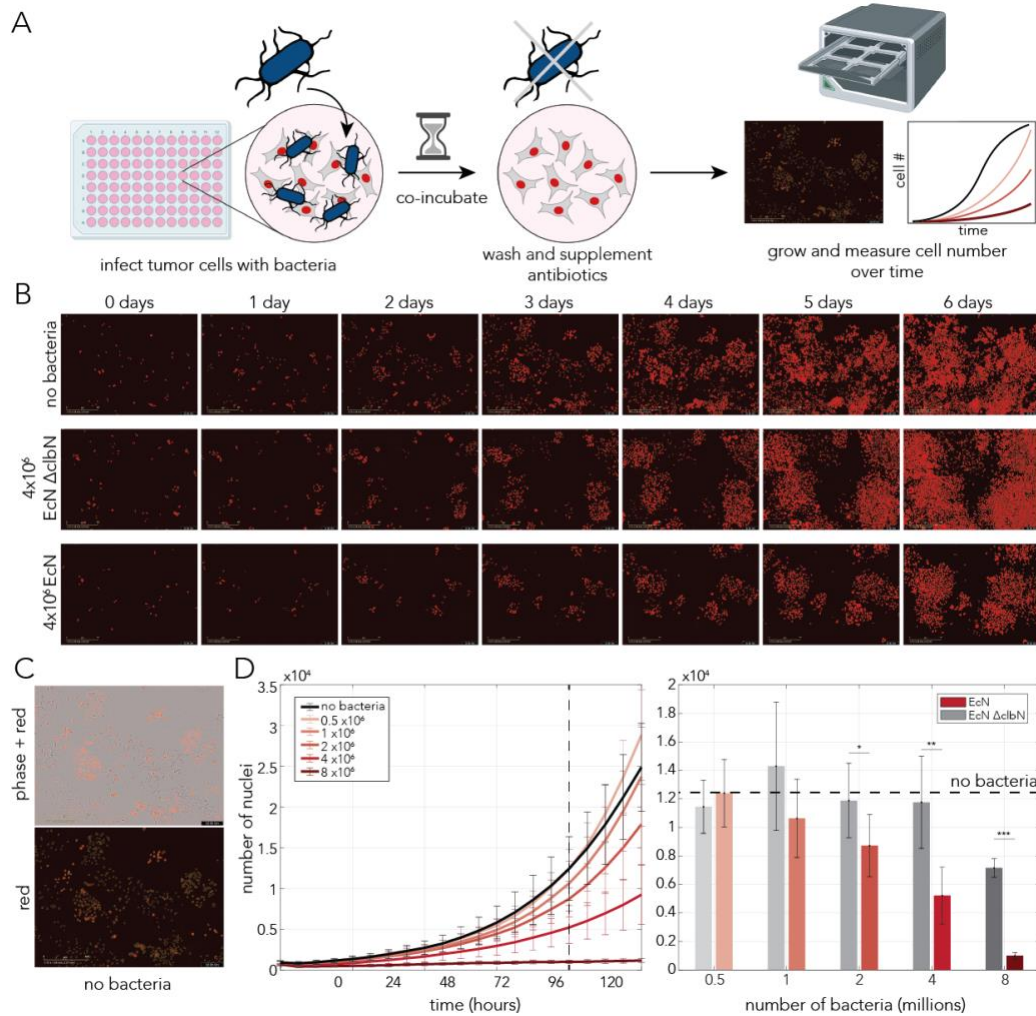


Figure 2.2 Colibactin, and not the presence of bacteria, induces cell cycle arrest.

(A) Schematic of experimental approach. HCT116 cells with a histone tag were co-cultured with bacteria. After the co-culture, bacteria were washed off and the tumor cells were cultured in media supplemented with antibiotics. Tumor cell proliferation was monitored with timelapse microscopy. (B) Representative images of three conditions every day over six days. The microscopy images are show the red fluorescence channel with a red mask segmenting the nuclei overlaid on the images. (C) representative images showing phase + red channel and red channel with nuclei outlines in green. These demonstrate how the cell nuclei were segmented independently of cell confluence. (D) Left panel shows growth curves of HCT116 cells co-cultured with different loads of EcN (shades of red) plotted with a no bacteria growth curve (black). The black vertical dashed line marks four days of growth at which point the number of nuclei per condition was determined and plotted in the right panel. EcN is again shown in shades of red by bacteria load and EcN Δ clbN is shown in shades of gray by bacteria load. The black horizontal dashed line marks the number of cells in the no bacteria control at four days. Error bars represent standard deviation. Each bacteria condition had four

replicates. No bacteria controls had 14 replicates. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

We next quantified the number of cells for each condition over time to determine the impact of colibactin on proliferation and if cells are able to recover at certain bacteria loads. We segmented the cell nuclei on the red channel using the histone tag and calculated the number of nuclei per well. Figure 2.2C shows representative images of phase and red overlaid and the red channel alone with outlines of the nucleus masks. The number of nuclei is plotted as growth curves of tumor cells averaged across replicates for cells exposed to different loads of EcN in Figure 2.2D (left panel). The bacteria loads of $0.5-8 \times 10^6$ encompassed a spectrum of proliferation phenotypes, ranging from no growth arrest to complete growth arrest. These curves can be compared to the no bacteria control, which shows that both the 0.5×10^6 and 1×10^6 appear to have little to no impact on cell proliferation. To quantify the bacteria loads inducing significant growth arrest, we compared the number of nuclei for each EcN and EcN $\Delta clbN$ bacteria load at the four-day time point (Figure 2.2D, right panel). The three highest bacteria loads significantly induced growth arrest in tumor cells exposed to colibactin. Only the highest bacteria load of 8×10^6 EcN $\Delta clbN$ also reduced cell proliferation, suggesting that proliferation is less sensitive than DNA damage to the number of bacteria in a four-hour co-culture. Taken together, these results establish the dose range for bacteria loads and time scales that induce detectable, yet recoverable, growth arrest. These doses and time scales can be applied to a genetic screen to further understand underlying mechanisms of colibactin toxicity.

2.3.3 Genetic screen uncovers the DNA damage response to colibactin

After we established ranges of bacteria at which we see reduced proliferation, but eventual cell growth, we applied this approach to a genetic screen to determine the cellular response to colibactin exposure. We opted to conduct a whole-genome CRISPR screen to systematically and unbiasedly map the DNA damage response to colibactin. Figure 2.3A outlines the overall approach of this screen. We introduced the Toronto knockout (TKO) library, containing nearly 71,000 single guide RNAs (sgRNAs) targeting over 18,000 genes, to HCT116 cells stably expressing Cas9. After establishing the library and expanding it, we conducted the screen at three bacteria loads: 2.26×10^9 (high), 1.13×10^9 (mid), and 0.565×10^9 (low). These bacteria loads are the equivalent of 4×10^6 , 2×10^6 , and 1×10^6 from the 96-well plate assays in Figure 2.2, but were scaled up to co-cultures in T175 flasks. These three bacteria loads were chosen as they all showed at least some growth inhibition followed by recovery, which is essential for detecting differential read counts of sgRNAs in a genetic screen. The screen co-cultures lasted four hours before tumor cells were washed and media, supplemented with antibiotics, was replenished. Each condition was allowed to recover until the flasks reached ~80% confluence, at which point cells were collected for DNA extraction and sgRNA sequencing. The high EcN condition did not recover after nine days and was excluded from sequencing.

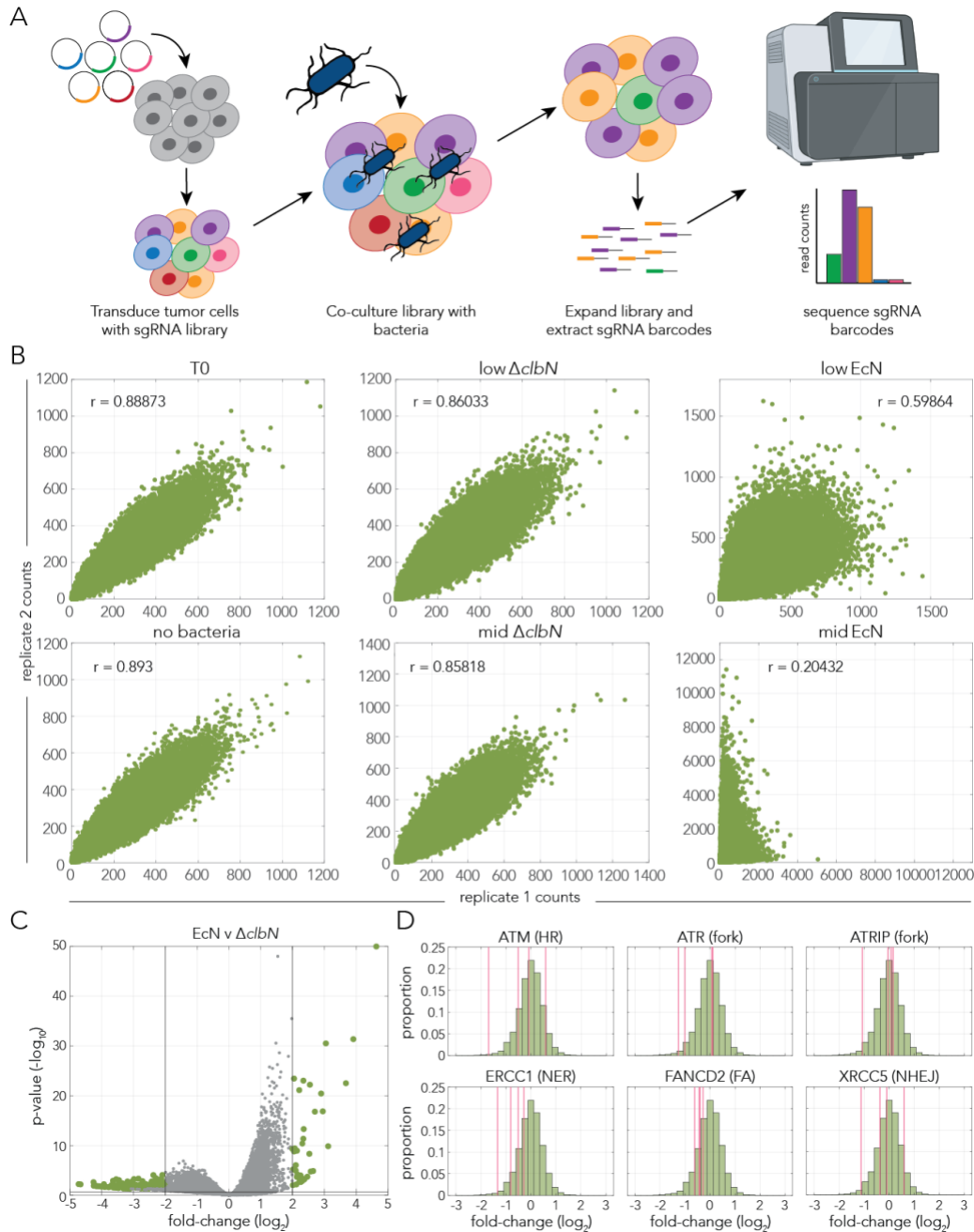


Figure 2.3 Genetic screen validates involvement of DNA repair genes in colibactin response.

(A) Overview of the whole-genome CRISPR screen in HCT116 cells. The TKO library was introduced to HCT116 cells. The library was co-cultured with bacteria and then allowed to recover and expand, at which point cells were collected for DNA extraction. The sgRNAs were sequenced and read counts determined for each guide. (B) Correlation plots of the normalized read counts for each sgRNA of two (out of three) replicates for each sequenced condition. (C) Volcano plot of the low EcN condition compared to the low EcN $\Delta clbN$ condition. sgRNAs

meeting both $\log_2(\text{fold-change})$ and p-value cutoffs are colored in green, all others are in gray. The vertical lines mark the fold-change cutoff of 4-fold and the horizontal line marks the p-value cutoff of 0.05. (D) Histograms of all sgRNAs (green) with vertical lines marking the fold-change of each of the four sgRNAs targeting specific genes marked in pink. The targeted pathways are in parentheses beside each targeted gene.

We first performed quality checks to confirm that our screen sequencing was sensible. Correlation plots comparing sgRNA normalized read counts of two (out of three) replicates revealed strong correlation of replicates for the initial library (T0), the no bacteria control, and both of the sequenced EcN $\Delta clbN$ conditions. The wild-type EcN conditions showed less correlation between replicates. The mid EcN condition showed very weak correlation between the two replicates and was excluded from further analysis. The mid EcN $\Delta clbN$ condition was also excluded as there was no relevant bacteria load to make comparisons with. We next evaluated differences in sgRNA counts between the low EcN and EcN $\Delta clbN$ conditions to identify enriched and depleted genes in colibactin-exposed tumor cells. While we were able to detect enrichment and depletion by fold-change, we observed weak p-values, seen as a flattened volcano plot in Figure 2.3C. To determine if we were able to detect underlying biology despite the noise in our screen, we extracted the fold-change of each sgRNA for genes previously shown to have a role in the DNA damage response to colibactin (Figure 2.3D). While there was variation between the fold-change of each sgRNA per gene, the general trend was that these knockouts were more sensitive to colibactin exposure, which agrees with previous observations (Sogari et al. 2024; Cuevas-Ramos et al. 2010; Bossuet-Greif et al. 2018; Dougherty et al. 2023).

Due to the variation in sgRNA responses per gene, we developed an alternative analysis approach to decrease the noise in our data. For each gene, we removed the two sgRNAs with the most extreme fold-change (absolute value) and took an average of the remaining two sgRNA fold-changes. We used a set of more than 800 olfactory receptor genes as a control set and fit a distribution curve to the observed fold-changes of these genes. We used this control distribution to determine empirical p-values for the genes targeted in this screen (Supplementary File 1). This approach likely increased the frequency of false negatives due to the likelihood of reducing the average fold-change for a gene below our cutoffs.

After reducing the noise in our dataset, we conducted a gene set enrichment analysis to determine which pathways are involved in responding to colibactin toxicity. Figure 2.4A shows the enrichment results of the Molecular Signatures Database Hallmarks gene set (results reported in Supplementary File 2). The most significant gene sets relate to cell cycle progression/proliferation and DNA damage. This aligns with what is already known about colibactin inducing cell cycle arrest and DNA damage. We further determined enrichment of DNA repair pathways defined by Olivieri et al. (2020) (results reported in Supplementary File 3). Colibactin-exposed cells were enriched for the FA/ICL, fork quality control, HR, NER, and EJ pathways, with FA/ICL and fork QC showing the strongest enrichment (Figure 2.4B). As HCT116 is MMR deficient, we did not expect to find enrichment for that pathway. However, the pathway likely does not play a

role in resolving colibactin-induced damage, as MMR typically repairs endogenously generated mutations. Interestingly, genetic screens conducted using other alkylating agents found that FA/ICL repair was the predominant pathway involved in repairing the DNA damage (Olivieri et al. 2020). The exception is cisplatin, where HR also played a key role. Thus, our data suggest that while colibactin alkylates DNA causing interstrand crosslinks, there may be more complexities to its interaction with DNA that involve the activation of nearly all repair pathways. These experiments demonstrate that co-culturing a knockout library with extracellular bacteria can lead to noisy results, but that there is still detectable underlying biology. We found that knockouts targeting the cell cycle and DNA repair sensitized cells to colibactin. When we further dissected the DNA repair pathways involved, we found that nearly all DNA repair pathways are activated by colibactin-induced damage, but that FA/ICL and fork QC appear to be the major operators.

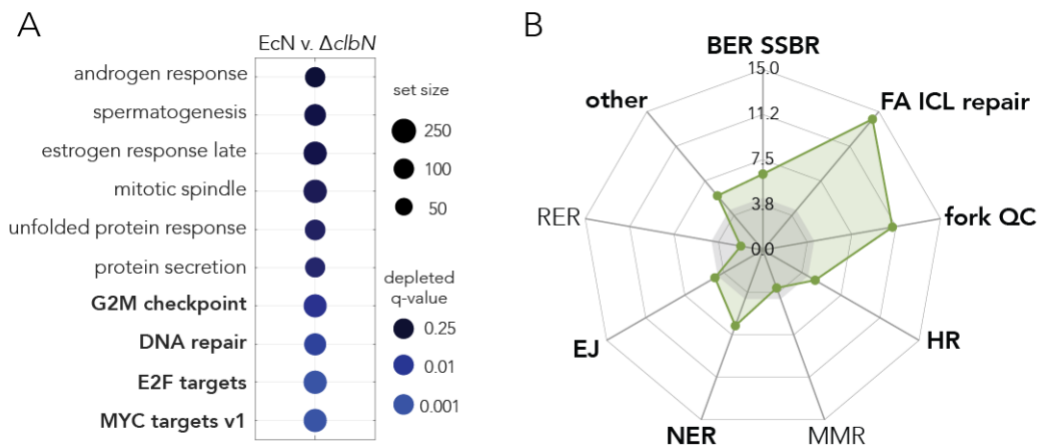


Figure 2.4 Colibactin-induced damage activates checkpoints and multiple DNA repair pathways.

(A) Enrichment plot for depleted pathways in the Molecular Signatures Database Hallmarks gene set (Liberzon et al. 2015). Circle size represents the gene set size

and shade of blue represents the q-value of these depleted gene sets. DNA damage and cell cycle gene sets are bolded. (B) Spider plot showing enrichment of DNA repair pathways as defined by Olivieri et al. (2020). Enrichment was determined by a Fisher's exact test between low EcN and low EcN $\Delta clbN$ ($-\log_2(\text{p-value})$ is plotted and labeled on the axes). Pathways are previously defined in the text with the exception of single strand break repair (SSBR) and ribonucleotide excision repair (RER). The significant pathways are bolded. The gray shaded region represents the significance level ($p = 0.05$).

2.4 Discussion

Colibactin was discovered nearly two decades ago as a DNA damaging agent, yet until now, a systematic approach to identify the DNA damage response has not been taken. This knowledge is especially important given the role that colibactin plays in tumorigenesis and its prevalence in some disease contexts. Several studies targeted a handful of genes in different pathways to propose the involvement of individual pathways in repairing damage from colibactin (Sogari et al. 2024; Cuevas-Ramos et al. 2010; Bossuet-Greif et al. 2018; Dougherty et al. 2023). At the time this screen was conducted, only non-homologous end joining and Fanconi anemia had been implicated in repairing colibactin damage (Cuevas-Ramos et al. 2010; Bossuet-Greif et al. 2018). A recent study identified an additional role for homologous recombination (Dougherty et al. 2023). However, two very recent studies took more systematic approaches, similar to our genetic screen, and were able to identify involvement of more repair pathways. Sogari et al. (2024) used a panel of CRC cell lines with mutations in different repair pathways to predict key pathways responding to colibactin-induced damage. This method identified additional involvement of nucleotide excision repair and replication fork quality control. Another more unbiased approach studied proteins

bound to chromatin with mass spectrometry. The chromatin was studied after conducting DNA replication reactions on plasmid DNA with colibactin crosslinks, which identified a role for translesion synthesis, Fanconi anemia, and homologous recombination in repairing the lesions (Altshuller et al. 2024).

This study is the first unbiased and systematic approach, using a whole-genome genetic screen, to characterize the DNA damage response following colibactin exposure. We first evaluated DNA damage as a reporter for colibactin toxicity, and while we observed significantly higher damage in colibactin-exposed cells, there were also high levels of damage from the presence of bacteria alone. This prompted us to determine if proliferation was a unique reporter for colibactin toxicity compared to the presence of bacteria. We established a range of bacteria loads that induced detectable growth arrest, but that had a subset of cells either recover or continue growing. These assays provided the conditions for conducting our genetic screen.

Our genome-wide approach, applied for the first time, is an unbiased method for fully mapping the cellular response in host cells to colibactin exposure. We created a DNA damage response profile (Figure 2.4B) showing that nearly all DNA repair pathways are activated by colibactin damage, particularly FA/ICL repair and fork quality control. The strength of involvement of these pathways likely relates to the detection of colibactin-induced interstrand crosslinks during replication, causing the replication fork to stall. This stalling would activate both

the FA/ICL pathway and the fork QC pathway to resolve the ICL and continue DNA replication. The role of the fork QC pathway appears to be unique to colibactin compared to other alkylating agents (Olivieri et al. 2020). Despite recent works establishing the involvement of many pathways in DNA repair following colibactin exposure, our screen is the first to identify a role for base excision repair/single strand break repair. Our screen validated the involvement of all other previously identified pathways, with the exception of translesion synthesis. TLS genes are classified in ‘other’ in Figure 2.4B, but only one (of the five) was a hit in our screen (REV3 was sensitive). However, we may be missing these genes as hits due to our analysis approach having a bias towards false negatives.

While we were able to detect involvement of specific DNA repair pathways, the overall results of the genetic screen were noisier than expected. Many CRISPR screens have been conducted with intracellular bacteria, but there is limited data from screens co-culturing host cells with extracellular bacteria, as we did. The number of bacteria needed to exert a detectable phenotype for bacteria-secreted products is high enough that there are adverse effects in the tumor cells from the bacteria alone. One approach to address this is to overexpress toxins like colibactin on bacterial artificial chromosomes, which may reduce the number of bacteria needed in the co-culture. In fact, co-cultures measuring γ -H2A.X following co-cultures with synthetically expressed colibactin observed minimal DNA damage when bacteria expressed only an empty chromosome (Nougayrède

et al. 2006; Secher et al. 2013; Bossuet-Greif et al. 2018; Sogari et al. 2024). A more thorough profile of the cellular response to colibactin could be obtained by repeating this genetic screen approach with a colibactin overexpression strain using a bacterial artificial chromosome in a non-cancerous cell line to make sure that all DNA repair pathways are fully functioning. This approach may also reveal how colibactin enters host cells, as our screen did not detect any clear mechanisms of entry. Despite these limitations in the current study, these results demonstrate that colibactin works uniquely in how it damages host DNA. Further structural analysis and genetic-based approaches may shed light on why so many DNA repair pathways respond to colibactin-induced damage. A better understanding of how colibactin damages host DNA and the resulting cellular response could also inform therapeutic targets for protecting host cells from colibactin toxicity.

2.5 Methods

Table 2.1 Bacteria strains used in this study

Strain	Nickname	Source
Nissle 1917	EcN	ArdeyPharm GmbH, (Pharma Zentrale GmbH), Germany
Nissle 1917 $\Delta clbN::chl$	EcN $\Delta clbN$	This study

Table 2.2 Cell lines used in this study

Cell line	Source
HCT116	ATCC
HCT116-H2B-mCherry	Produced by lentiviral transduction by a previous lab member

HCT116-cas9	This study
HCT116-cas9-TKO	This study

2.5.1 Cell lines and media

Cell lines were cultured in DMEM supplemented with 5% fetal bovine serum (Gibco FBS cat# 26140-079) and 25 mM HEPES Buffer (Corning, Cat# 25-060CI) and incubated at 37 °C with 5% CO₂. All experiments were conducted by plating cells in 96-well plates or T-175 flasks 16–18 h before starting the experiment. Cells were passaged at 70-80% confluence. Bacteria cultures were grown in Lysogeny Broth (LB) at 37°C with 200 rpm orbital shaking. Bacteria strains with antibiotic resistance were grown with 25 µg/mL chloramphenicol.

2.5.2 Cloning deletion strains

clbN was deleted from Nissle 1917 using lambda red recombination with an insert containing 40 bp homology arms to the upstream and downstream genomic regions, respectively, and a chloramphenicol resistance cassette.

2.5.3 Co-culture assays in 96-well format

HCT116 cells were seeded at 10,000 cells per well in a 96-well-plate (Eppendorf, #0030730119) in 100 µL DMEM. Overnight cultures of EcN and EcN $\Delta clbN$ were grown in 3 mL LB supplemented with the relevant antibiotics. 1 mL of overnight cultures was washed twice in PBS before resuspending in DMEM. Overnight cultures were diluted 1:20 and incubated for one hour at 37°C with 200 rpm orbital shaking. We measured the OD₆₀₀ and normalized both strains to an

OD₆₀₀ of 0.1. This stock was serially diluted 1:1 in DMEM to make seven dilutions and one no bacteria control. 100 µL of each dilution was added to three replicate wells in the 96-well plate with HCT116 cells pre-seeded. This resulted in a final volume of 200 µL media per well and three replicates per condition. The co-culture plate was incubated at 37°C with 5% CO₂. After four hours of co-culture, 180 µL of media was aspirated from wells and 180 µL of PBS added. This step was repeated for twice. After the final PBS wash, plates were processed for immunofluorescence staining or 200 µL DMEM supplemented with 50 µg/mL gentamicin was replaced and the HCT116 cells allowed to grow and kill any remaining bacteria.

2.5.4 Immunofluorescence staining

Following the bacteria wash steps after a four-hour co-culture, cells were fixed in 3.7% formaldehyde at room temperature for 20 minutes. The wells were washed three times with PBS to remove any remaining formaldehyde and terminate the fixation step. Cells were permeabilized with PBS supplemented with 0.5% Triton X-100 (Sigma, #X100) at room temperature and 300 rpm orbital shaking for five minutes. The permeabilization buffer was removed and 3% BSA solution (Research Products International Corp, #B41600) was added to each well to block cells for one and a half hours at room temperature with 300 rpm orbital shaking. The blocking buffer was removed and 40 µL of p-histone H2A.X (Cell Signaling, #9718T) antibody diluted 1:800 in 3% BSA solution was added to each well. Cells were incubated overnight at 4°C with rocking. Wells were then washed five

times with 200 μ L PBS for five minutes each at room temperature with 300 rpm orbital shaking. 40 μ L of the secondary goat anti rabbit antibody (Invitrogen, #A21244, Alexa Fluor 647), 4 μ g/mL in 3% BSA blocking buffer, was added to each well and incubated at room temperature for one hour with 300 rpm orbital shaking and protected from light. The antibody was removed and wells were washed two times with 200 μ L PBS for five minutes at room temperature with 300 rpm orbital shaking and protected from light. Wells were stained with 50 μ L DAPI (Invitrogen, D1306) diluted to 300 nM in PBS for 10 minutes at room temperature and protected from light. Wells were washed three times with 200 μ L PBS for five minutes at room temperature and 300 rpm orbital shaking and protected from light. After the final wash, 100 μ L of PBS was added to wells to prevent the cells from drying out and the plate was stored in foil to protect from light at 4°C until imaging.

2.5.5 Immunofluorescence microscopy and analysis

Microcopy of immunofluorescently stained cells was performed on a Zeiss Axio Observer.Z1 epifluorescence microscope. The DAPI (465 nm), Alexa Fluor 647 (668 nm) and brightfield channels were used at 20x magnification. Six images were taken per well.

Image analysis was conducted with custom MATLAB (MathWorks) scripts. Briefly, cell nuclei were segmented with an automatic threshold on the DAPI channel. Masks were excluded if they were too small (debris) or too large

(multiple nuclei overlapping). The AF647 (pH2A.X) mean intensity was determined within each mask per image and concatenated to all other measurements from that well (six images total per well).

2.5.6 Proliferation assays

96-well plates were imaged on an IncuCyte S3 automated microscope with the red and phase channels immediately before bacteria were added to get the initial number of HCT116 cells. Imaging was scheduled for every six hours, thus the second time point was after the bacteria were washed from the wells and gentamicin was added to kill any remaining bacteria. Plates were imaged every six hours for up to six days.

2.5.7 TKO library generation

The Toronto knockout library was a gift from Jason Moffat (Addgene #125517) (Mair et al. 2019). Lentivirus was produced by transfecting either the lentiCas9-Blast (Addgene #52962) or TKOv3 plasmids with vectors encoding packaging proteins (pMD2.G and pCMVdr8.91) with lipofectamine 3000 (Invitrogen, L3000001) into HEK-293T cells seeded at ~30% confluency. Viral supernatant was collected, filtered with a 0.22 μm filter, and frozen for future use. HCT116 cells were transduced with the lentiCas9-Blast virus and maintained under blasticidin (Invitrogen, #R210-0110, $\mu\text{g}/\text{mL}$) selection. HCT116-cas9 cells were transduced with the TKOv3 lentivirus at an MOI of ~0.3 and coverage of 500x per guide. Cells were selected with puromycin (1 $\mu\text{g}/\text{mL}$) and frozen maintaining at least 400x coverage.

2.5.8 CRISPR screen

Approximately 18 hours before the screen, T175 flasks (Eppendorf, 50995801) were seeded with 8×10^6 million HCT116-cas9-TKO cells. We seeded two flasks per replicate and three replicates per bacteria dose for a started coverage of $\sim 225x$. We also seeded three no bacteria control replicates and three T0 replicates (with three flasks each for $\sim 330x$ coverage for an accurate starting library). Overnight cultures of EcN and EcN $\Delta clbN$ were grown in 25 mL LB supplemented with the relevant antibiotics for ~ 16 hours. The following day, 10 mL of each overnight culture was pelleted ($3000xg$, 10 minutes) and washed twice in PBS. We measured OD_{600} of each strain three times and took the average of all three measurements. The volume needed for the high dose (2.26×10^9 per flask, or $4.52 \times 10^8/mL$) was pelleted and resuspended in 50 mL cell culture media (DMEM) for each strain. This stock was diluted 1:1 two times to get the mid and low bacteria loads. We added five mL of each dilution per strain to six flasks. The no bacteria flasks received five mL plain media. The co-cultures were then incubated for four hours. Media containing bacteria was removed and the tumor cells were washed three times in PBS. Media was replenished and supplemented with $50 \mu g/mL$ gentamicin. The following day, tumor cells were washed again two times with PBS to remove any dead bacteria and media (supplemented with $50 \mu g/mL$ gentamicin) was replenished.

T0 cells were collected at the start of the co-culture and frozen as pellets of 24×10^6 cells (~330x coverage) for later sequencing. Every 72 hours cells were detached with TrypLE (Gibco, 12605010) to promote an even layer of cell growth and refresh the media. Cells were maintained on media supplemented with 50 $\mu\text{g}/\text{mL}$ gentamicin for the duration of the experiment. All other conditions were collected and pelleted (24×10^6 cells) when flasks reached ~80% confluence. The high EcN treated tumor cells never recovered and thus were not collected for sequencing.

2.5.9 CRISPR screen sequencing

DNA was extracted from cell pellets with the Qiagen Blood and Cell Culture DNA Maxi kit (Qiagen, 13362). Precipitated DNA was dissolved in molecular grade water. 30-40 PCR reactions were set up per DNA sample to maximize coverage of each sample and achieve at least 200x coverage. Each PCR reaction had 3.5 μg of DNA and was performed with NEBNext Ultra II Q5 Master Mix (NEB, #M0544). Primers used in the PCR are listed below and the cycles are in Table 2.3. A one-step PCR was performed using primers with both sequence homology to amplify the gRNA and Illumina adaptors and indexes for sequencing.

Example primer 1:

CAAGCAGAAGACGGCATAACGAGATXXXXXXXXXXGTGACTGGAGTTCAG
ACGTGTGCTCTTCCGATCT(N)ACTTGCTATTCTAGCTCTAAAAC

(Illumina i7 index (X), 0-3 random nucleotide stagger (N), gRNA flanking
homology)

Example primer 2:

AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXACACTCTTTC

CCTACACGACGCTCTTCCGATCT(N)GAGGGCCTATTTCCCATGATTC

(Illumina i5 index (X), 0-3 random nucleotide stagger (N), gRNA flanking
homology)

Table 2.3 One-step PCR protocol

98°C	30 sec	
98°C	10 sec	x27 cycles
61°C	30 sec	
72°C	20 sec	
72°C	2 min	
4°C	Hold	

Following the PCR, all reactions per sample were pooled and 100 µL was run on a 2% agarose gel. A product of ~400 bp was excised from the gel and purified with ZR-96 Zymoclean Gel DNA Recovery Kit (Zymo, #D4022). Library quality was assessed on a BioAnalyzer with the Agilent High Sensitivity DNA Kit (Agilent Technologies, Cat# 5067-4626). Samples were normalized to 10 nM and pooled for sequencing. Libraries were denatured and diluted according to the NextSeq 500/550 system protocol and sequenced with the NextSeq 500/550 High Output kit v2.5, 75 cycles (Illumina, cat# 20024906) on a NextSeq 500/550 machine.

2.5.10 CRISPR screen analysis

We trimmed fastq files with trimmomatic (Bolger et al. 2014) and performed sequence alignment with bowtie (Langmead et al. 2009). We then used MAGeCK (Li et al. 2014) to generate read counts and normalized counts. We wrote a custom MATLAB (MathWorks) script to determine empirical p-values for each gene to minimize noise in our sequencing data. Briefly, we excluded the two largest fold-changes of sgRNAs per gene (absolute value) and took the average of the remaining two sgRNAs as the fold-change for each gene. We then fit a distribution to over 800 olfactory receptor sgRNAs and determined empirical p-values based on the frequency each fold-change was observed. We conducted a gene set enrichment analysis (Subramanian et al. 2005; Mootha et al. 2003) with the Molecular Signatures Database Hallmark Gene Set (Liberzon et al. 2015) to identify enriched pathways.

2.5.11 Data and code availability

All raw sequencing data from the CRISPR screen is available in the NCBI BioProject database under accession number [XYZ](#). Code for all analyses can be found on GitHub (DOI [XYZ](#)).

Chapter 3: Colibactin leads to a bacteria-specific mutation pattern and self-inflicted DNA damage

3.1 Preface

This chapter is adapted from a manuscript currently under review. The authors on this manuscript are Emily Lowry, Yiqing Wang, Tal Dagan, and Amir Mitchell.

The author contributions are as follows:

Conceptualization: A.M. and E.L. Methodology: A.M. and E.L. Investigation: E.L. Visualization: A.M. and E.L. Formal analysis: A.M., E.L., Y.W., and T.D. Funding acquisition: A.M. and T.D. Project administration: A.M. Supervision: A.M. Writing – original draft: A.M. and E.L. Writing – editing: A.M. and E.L. Writing – review: A.M., E.L., Y.W., and T.D.

Yiqing Wang and Tal Dagan contributed to the data analysis of the mutation accumulation experiment and the results presented in Figure 3.3F. I performed all experiments and conducted all other data analysis myself or with the assistance of Amir Mitchell.

3.2 Abstract

Colibactin produced primarily by *Escherichia coli* strains of the B2 phylogroup crosslinks DNA and can promote colon cancer in human hosts. We investigated the toxin's impact on colibactin producers and on bacteria co-cultured with producing cells. Using genome-wide genetic screens and mutation accumulation experiments we uncovered the cellular pathways that mitigate colibactin damage and revealed the specific mutations it induces. We discovered that while

colibactin targets A/T rich motifs, as observed in human colon cells, it induces a bacteria-unique mutation pattern. Based on this pattern, we predicted that long-term colibactin exposure will culminate in a genomic bias in trinucleotide composition. We tested this prediction by analyzing thousands of *E. coli* genomes and found that colibactin-producing strains indeed show the predicted skewness in trinucleotide composition. Our work revealed a bacteria-specific mutation pattern and suggests that the resistance protein encoded on the colibactin pathogenicity island is insufficient in preventing self-inflicted DNA damage.

3.3 Introduction

Competitive interactions are common in microbial communities, including the human gut microbiome (Kern et al. 2021). Secreted toxins that target neighboring microbes are a common mechanism underlying such competitive interactions. Colibactin is a bacteria-secreted genotoxin that can bind DNA in neighboring cells (Vizcaino and Crawford 2015; Pleguezuelos-Manzano et al. 2020; Nougayrède et al. 2006). Colibactin damage is toxic for some bacteria species (Silpe et al. 2022; Wong et al. 2022; Chen et al. 2022) and can also harm the host intestinal cells (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). Colibactin-producing bacteria have been associated with multiple human diseases ranging from inflammatory bowel disease to colorectal cancers (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). While colibactin-induced damage in host cells is extensively investigated due to its clinical relevance, its toxicity to bacteria remains underexplored. Addressing this

knowledge gap could elucidate how colibactin impacts the host microbiome and whether colibactin expression is associated with any burden on cells producing the toxin. We examined multiple aspects of colibactin toxicity in *Escherichia coli*. We discovered that while colibactin acts as a mutagen and thus depends on a similar mechanism of action in both bacterial and mammalian systems, fundamental differences in bacteria exist.

Colibactin is encoded by a 54 kb genomic region known as the *pks* island. This region comprises 19 genes needed to synthesize and export the toxin, including non-ribosomal peptide synthetases and polyketide synthases (Nougayrède et al. 2006). The island also encodes a cyclopropane hydrolase (*clbS*) that protects colibactin-producing cells from the toxin (Nougayrède et al. 2006; Bossuet-Greif et al. 2016; Tripathi et al. 2017). The toxin itself contains two cyclopropane rings that alkylate DNA and cause interstrand crosslinks (Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). Colibactin is most commonly found in *E. coli* strains of the B2 phylogroup and is expressed by both pathogenic and commensal strains (Nougayrède et al. 2006; Wami et al. 2021). Bacteria harboring the *pks* island are estimated to exist in the gut microbiome of 20-30% of healthy individuals (Nougayrède et al. 2006; Dubois et al. 2010; Watanabe et al. 2020; Dejea et al. 2018; Eklöf et al. 2017; Dubinsky et al. 2020). The prevalence of *pks* positive bacteria increases to approximately 60% in patients with colorectal cancer and

inflammatory bowel disease (Buc et al. 2013; Dejea et al. 2018; Eklöf et al. 2017; Dubinsky et al. 2020).

The clinical relevance of colibactin in various human diseases underlies the widespread efforts to study its toxicity in mammalian models. *In vitro* experiments showed that colibactin causes DNA damage and leads to cell-cycle arrest in various mammalian cells. This damage requires contact between bacteria and the targeted host cells (Bossuet-Greif et al. 2018; Nougayrède et al. 2006; Reuter et al. 2018; Silpe et al. 2022; Wong et al. 2022). Colibactin interstrand crosslinks are resolved through the activation of multiple DNA repair pathways, including the non-homologous end-joining (Cuevas-Ramos et al. 2010), homologous recombination (Dougherty et al. 2023), and the Fanconi anemia repair pathways (Bossuet-Greif et al. 2018; Dougherty et al. 2023). Since repair of DNA crosslinks introduces double-stranded breaks, they are potentially mutagenic, leading to the hypothesis that colibactin may predispose individuals to colon cancer (Cuevas-Ramos et al. 2010). In human cells, colibactin-induced mutations are commonly found in hexameric A/T-rich DNA motifs (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). This enrichment is attributed to the particularly narrow minor groove of these sequences which promotes colibactin binding (Dziubańska-Kusibab et al. 2020). Recent work has found that a colibactin-associated mutational signature is detected in 5-10% of colorectal cancers (Pleguezuelos-Manzano et al. 2020), supporting its involvement in cancer.

Colibactin toxicity has also been observed in bacteria. Auto-toxicity was observed in colibactin-producing bacteria genetically engineered to knockout *clbS* (the protective cyclopropane hydrolase) and became more pronounced upon inactivation of the nucleotide excision repair pathway (Tripathi et al. 2017; Bossuet-Greif et al. 2016). Studies also showed colibactin can target other bacterial species, including several *Staphylococcus* species (Wong et al. 2022; Chen et al. 2022), several *Vibrio* species, *Clostridium difficile*, and *Enterobacter aerogenes* (Chen et al. 2022). A recent study suggested that colibactin damage arises from prophage induction (Silpe et al. 2022). However, prophage-cured *S. aureus* remain susceptible to colibactin, indicating that toxicity mechanisms beyond prophage-induction exist (Wong et al. 2022). Despite the multiple reports of colibactin damage in bacteria, key gaps in knowledge persist. A key open question in bacteria includes which specific DNA repair pathways mitigate colibactin damage. Given that DNA is fundamentally differently packaged in eukaryotes and bacteria, it remains unknown whether colibactin favors A/T rich regions in bacteria and if it culminates in similar mutations as those reported for colon cells.

Here, we aimed to identify the environmental conditions that maximize colibactin toxicity and uncover genes that modulate its toxicity using a genome-wide loss-of-function genetic screen. We then investigated if colibactin induces genetic mutations in *E. coli* cells that are co-cultured with colibactin producing cells and

whether a colibactin-specific mutational signature can be detected. Finally, we evaluated colibactin production induces self-inflicted damage both *in vitro* and *in silico*.

3.4 Results

3.4.1 Colibactin reduces viability of co-cultured *E. coli*

We first determined if colibactin reduces viability in neighboring bacteria. We co-cultured the reporter strain (ampicillin-resistant lab strain) with colibactin-producers, *E. coli* cells from the same genetic background that harbor the pathogenicity island (pks^+) on a bacterial artificial chromosome (BAC) (Silpe et al. 2022). We grew cells as a pellet to maximize cell-cell contact for the duration of the co-culture experiment. Bacteria transformed with an empty BAC were used as a negative control (pks^-). Following co-culture, we calculated the colony forming units (CFUs) of the reporter strain by plating on selective agar. Figure 3.1A shows the experimental setup and representative plate images showing reduced viability. We measured viability over 48-hours and observed that colibactin reduces viability within 12 hours of co-culture and continues to decrease viability of co-cultured cells over time (Figure 3.1B). Since growth of the wild-type strain in the pellet is arrested after the 12-hour time point in the control experiment, reduced viability in the pks^+ condition is likely attributed to cell death and not just growth arrest.

We then wanted to quantify colibactin toxicity under different environmental conditions since previous work showed that colibactin expression is influenced by environmental conditions (Tronnet et al. 2016, 2017; Oliero et al. 2021; Chagneau et al. 2019; Bossuet et al. 2023). We compared co-cultures in different media, in pellet or suspension, and at different strain ratios. Figure 3.1C shows the results of these experiments. We observed that colibactin toxicity was higher in nutrient-poor media compared to nutrient-rich media and was also higher when the co-culture grew in a pellet compared to growth in suspension. Lastly, toxicity was increased when the co-culture ratio was skewed towards colibactin producers. Taken together, these experiments allowed us to uncover the experimental settings that maximize colibactin toxicity.

3.4.2 Genetic screen uncovers genes mitigating colibactin toxicity

In order to identify genes and pathways involved in mitigating colibactin-induced damage, we used a loss-of-function genetic screen that was based on results from our reduced viability assay (Figure 3.1C). We used a pooled genetic screening approach that we recently used for studying drug sensitivity (Noto Guillen et al. 2024). Briefly, a collection of 7,259 *E. coli* knockout strains targeting 3,680 non-essential genes were co-cultured with the colibactin-producing strain. Changes in strain frequency were deduced by sequencing unique DNA barcodes that identify each knockout strain. We conducted the screen with different levels of selective pressure (co-culture ratios and incubation durations). Co-cultures with a *pks*⁻ strain were used as controls. Overall, we conducted six screens with five

biological replicates each (see Methods). At the end of each screen, cells were collected for DNA extraction, barcode amplification, and DNA sequencing (Figure 3.1D). In these screens, barcode depletion in the pks^+ condition relative to the pks^- condition is indicative of a gene that mitigates colibactin-induced damage (Figure 3.1D).

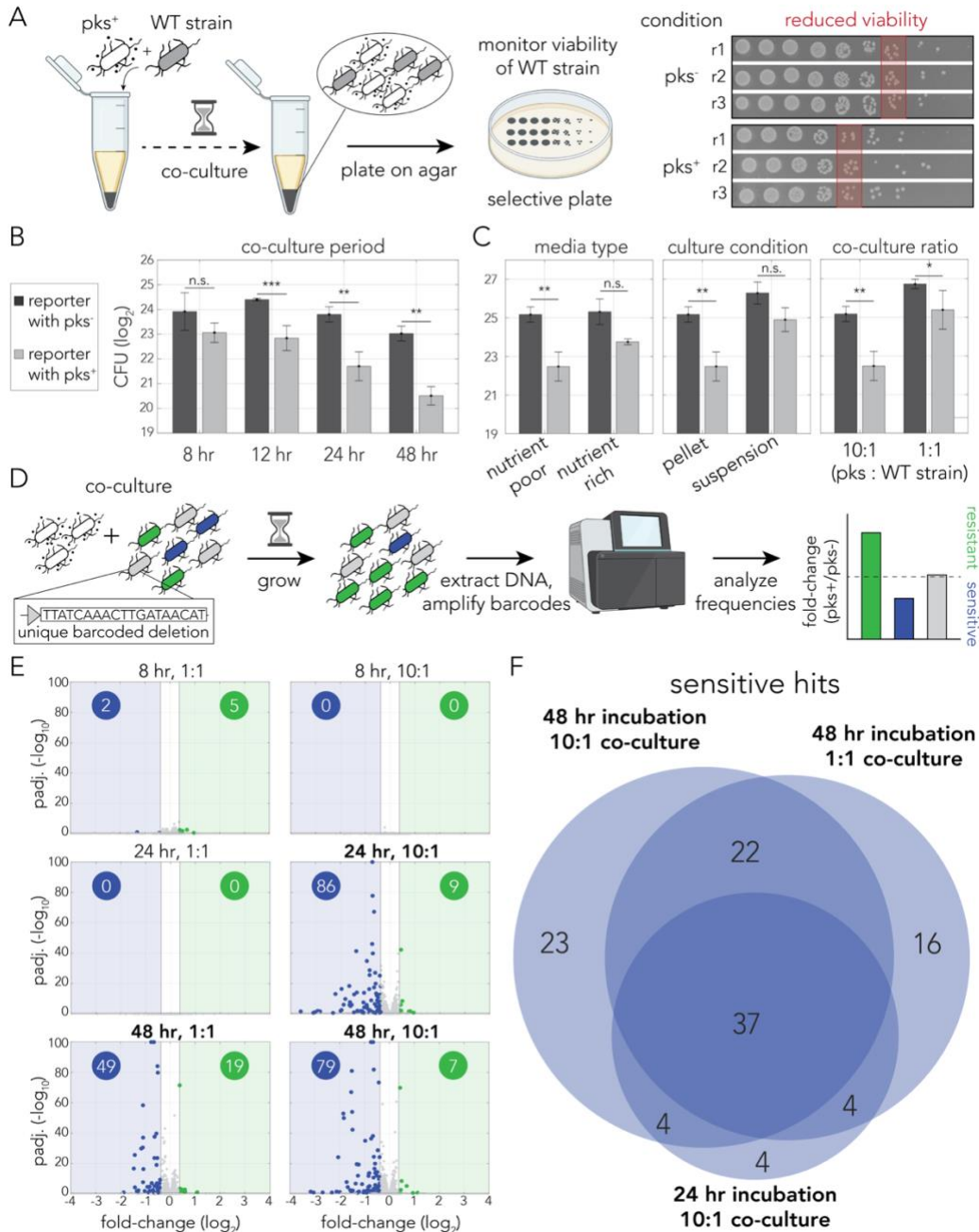


Figure 3.1 Genetic screen reveals key role of homologous recombination in colibactin response.

(A) Overview of the reduced viability assay in response to colibactin. Reporter cells were co-cultured with *pks*⁺ cells in a pellet and plated on selective agar to determine the number of viable reporter cells. The colony images on the right show representative results of a spotting assay with reduced viability after *pks*⁺ co-cultures. (B) Colibactin toxicity correlates with co-culture period. A reporter strain was co-cultured in pellets with *pks*⁺ (light gray) or *pks*⁻ (dark gray) strains at a 10:1 ratio in M9. The bar graphs show the mean CFUs back calculated from the spotting assay. Error bars show standard deviation of triplicates (***p*<0.01, ****p*<0.001, two sample *t*-test). (C) Colibactin toxicity is influenced by co-culture conditions. All conditions were evaluated at 24 hours. Toxicity is impacted by growth media (left panel): Co-cultures were conducted in either nutrient poor (M9) or rich media (LB) at a 10:1 ratio and pelleted. Toxicity is impacted by growth conditions (middle panel): Co-cultures were in nutrient poor media at a 10:1 ratio either pelleted or in suspension. Toxicity is impacted by incubation ratio (right panel): Co-cultures were in nutrient poor media at either a 10:1 or 1:1 ratio of *pks*⁺ to reporter strain and pelleted. Error bars show standard deviation of triplicates (**p*<0.05, ***p*<0.01, two sample *t*-test). (D) Overview of barcoded genetic screen approach. (E) Volcano plots of screen results. Resistant hits are colored in green and sensitive hits are colored in blue. The number of hits for each direction are reported in colored circles on each volcano plot. Vertical gray lines represent the fold-change cutoff. (F) Venn diagram of shared sensitive hits between 24-hour 10:1, 48-hour 1:1, and 48-hour 10:1 screens.

Analysis of the DNA sequencing results identified at least 2.6×10^6 barcode sequences in each experiment. 95% of individual barcodes were sequenced more than 10 times and the median depth per barcode was 1,161. Figure 3.1E shows volcano plots for each of the screening conditions. As expected, we detected more hits when the target bacteria had longer exposure to and were in the presence of more colibactin producers (Figure 3.1E). We identified a total of 110 knockout strains that conferred colibactin-sensitivity and 42 knockout strains that conferred resistance in at least one condition (Supplementary File 4). Reassuringly, when analyzing the overlap in the 110 sensitive hits found across the three most toxic condition (marked in bold in Figure 3.1E), we observed that 67 of them were

shared in at least two conditions (Figure 3.1F). We noted that contrary to screens we previously performed with this strain collection (Sayin et al. 2023; Rosener et al. 2020), we observed a high degree of variability between biological replicates, manifesting in a marginal p-value for some of the top hits (Figure 3.1E). We expect that this variability arises from cells randomly detaching from the pellet and growing without consistent colibactin exposure. As detached cells are also collected and used for DNA extraction, they are included in the analysis and can potentially give rise to false positive hits.

We next conducted a functional enrichment analysis of the genome-wide screen using KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto 2000) and GO terms (The Gene Ontology Consortium et al. 2000) to evaluate pathways involved in colibactin-induced toxicity (Supplementary File 5). This analysis identified pathways directly or indirectly connected to DNA synthesis and repair as conferring sensitivity (Figure 3.2A). These pathways include homologous recombination, purine metabolism, and the master pathway for amino sugar and nucleotide sugar metabolism, which is connected to the stringent response. Knockout of the flagellar assembly pathway also increased sensitivity, which may potentially be due to the inability of these strains to escape the co-culture pellet.

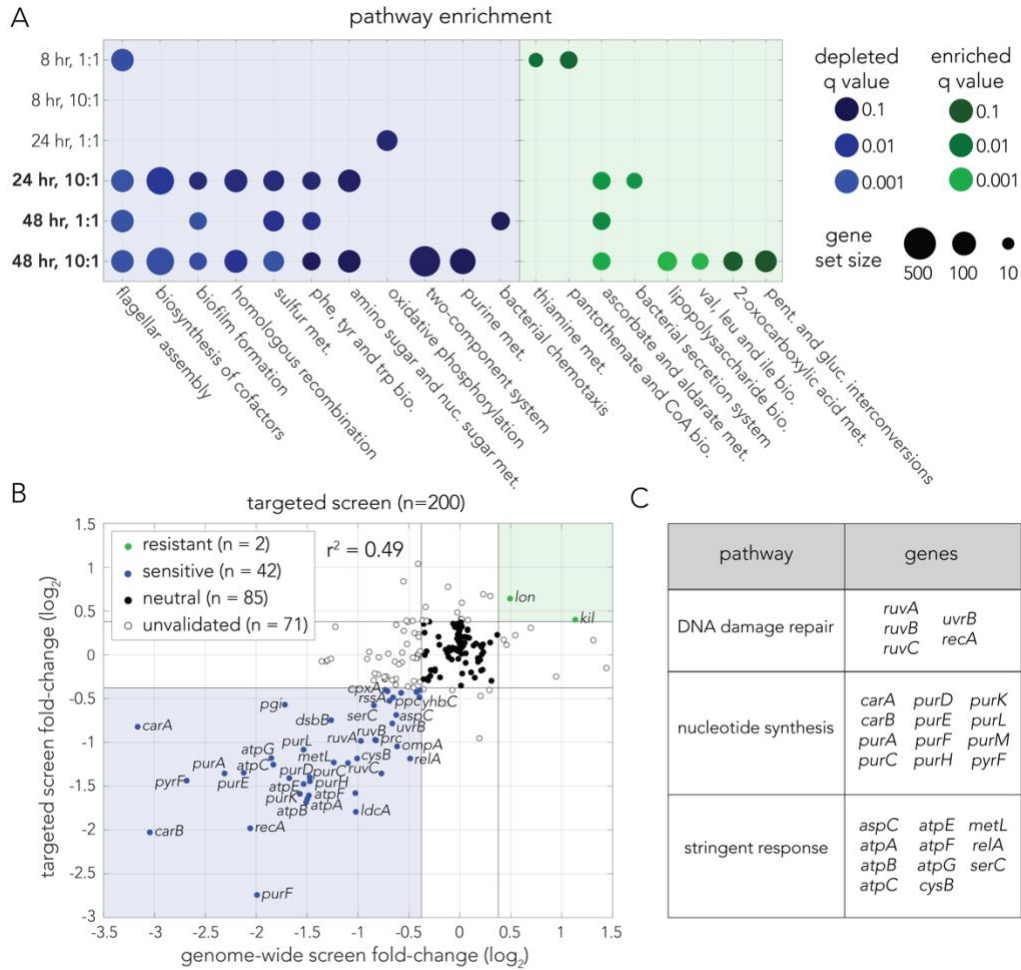


Figure 3.2 Pathway enrichment and targeted validation screen identify role of homologous recombination in colibactin response.

(A) Pathway enrichment analysis using KEGG (Kanehisa and Goto 2000) terms for each screen. Enriched terms are colored in green by q value and depleted terms are colored in blue by q value. Circle size represents the size of the gene set for the respective pathway. (B) Comparison of knockout fold-changes in the genome-wide screen and targeted validation screen. Points are colored by consensus enrichment change in the genome-wide screen and targeted screen (green for resistant, blue for sensitive, black for neutral, and empty for unvalidated). Gray lines mark the fold-change cutoff (1.3). (C) Key pathways underlying colibactin sensitivity. Twenty-eight of the validated colibactin-sensitive knockouts are associated with three cellular pathways.

Given the relatively high variability we observed between replicates in the genome-wide screen, we followed with a targeted validation screen. This screen

included all 162 original hits and 38 additional neutral knockout strains. We performed the screen with six replicates at the most extreme co-culture condition (colibactin:target bacteria at a 10:1 ratio for 48 hours). Figure 3.2B shows a comparison between the genome-wide and targeted screens ($r^2 = 0.49$, p -value $< 10^{-10}$ by Pearson's correlation test). Results for all 200 strains in the validation screen appear in Supplementary File 5. Inspection of the genes mitigating colibactin-induced damage revealed hits from three central pathways involved in DNA damage repair (Figure 3.2C). We found multiple genes associated with homologous recombination (*ruvABC* and *recA*) (Kowalczykowski et al. 1994) but only one from the nucleotide excision repair pathway (*uvrB*), which was previously linked to self-inflicted colibactin toxicity (Bossuet-Greif et al. 2016). We noted that knockouts from other DNA repair pathways were also observed to increase colibactin sensitivity, yet were below our cutoff value we chose for fold-change. These knockouts included *uvrA* of the nucleotide excision repair pathway and *umuCD* that encode polymerase V involved in translesion synthesis (Supplementary File 4). We also found twelve sensitive hits from the nucleotide synthesis gene network, likely increasing sensitivity to DNA damage by reducing nucleotide availability (Rosener et al. 2020). Lastly, eleven of the sensitive hits belong to the stringent response that works along with the SOS response to induce mutagenic DNA repair (Shee et al. 2011; Ponder et al. 2005; Zhai et al. 2023).

Taken together, our screens revealed that inactivation of multiple pathways increases colibactin sensitivity, and that they all likely operate by modulating the

DNA damage response. Moreover, within the DNA damage response we primarily detected genes involved in homologous recombination but only one gene linked to nucleotide excision repair.

3.4.3 Colibactin induces a specific mutational signature

Colibactin-induced damage leads to a specific mutational signature in colorectal cancers (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020; Chen et al. 2023) that is attributed to its favorable binding to the particularly narrow minor groove in A/T rich sequence motifs (Dziubańska-Kusibab et al. 2020). We sought to determine whether the mutational pattern is the same in bacteria using mutation accumulation experiments with repeated exposure to colibactin producers. Ampicillin-resistant DNA damage reporter cells were co-cultured for 24 hours with genetically engineered colibactin producers and spotted on selective agar. A single reporter colony was then grown and used for a subsequent exposure cycle. We repeated the experiment with 48 independent replicates exposed to colibactin (pks^+) and 48 controls (pks^-). After completing ten exposure cycles we sequenced the genome and annotated the mutations with the breseq tool (Deatherage and Barrick 2014) (Figure 3.3A). Since colonies were selected randomly in each cycle, observed mutations are not expected to be adaptive ones but should reflect the pattern of spontaneous mutation induced by colibactin exposure.

Figure 3B shows a summary of the observed mutations by category, which are also reported in Supplementary File 6. Overall, we observed considerably more mutations across all categories in the *pks*⁺ co-culture condition. We observed a ten-fold increase in single base substitutions (SBS) and only a two-fold increase in short indels from co-culture with *pks*⁺ cells (short indels are highly prevalent in colon cells (Pleguezuelos-Manzano et al. 2020; Chen et al. 2023)). To account for transcription coupled repair we also tested for mutational strand bias. We did not detect a SBS bias when comparing the transcribed and untranscribed strands. Lastly, we also observed a very high frequency of a single large deletion that was almost identical across 30 of the 48 *pks*⁺ co-culture replicates. The genome position of this deletion aligns with the known location of the cryptic prophage e14 (Figure 3.3C) and agrees with the known role of colibactin in inducing prophage excision (Silpe et al. 2022). A high frequency of e14 excision relative to other cryptic prophages was also observed after exposure to the mitomycin-C alkylating agent which also causes DNA damage (Wang et al. 2010).

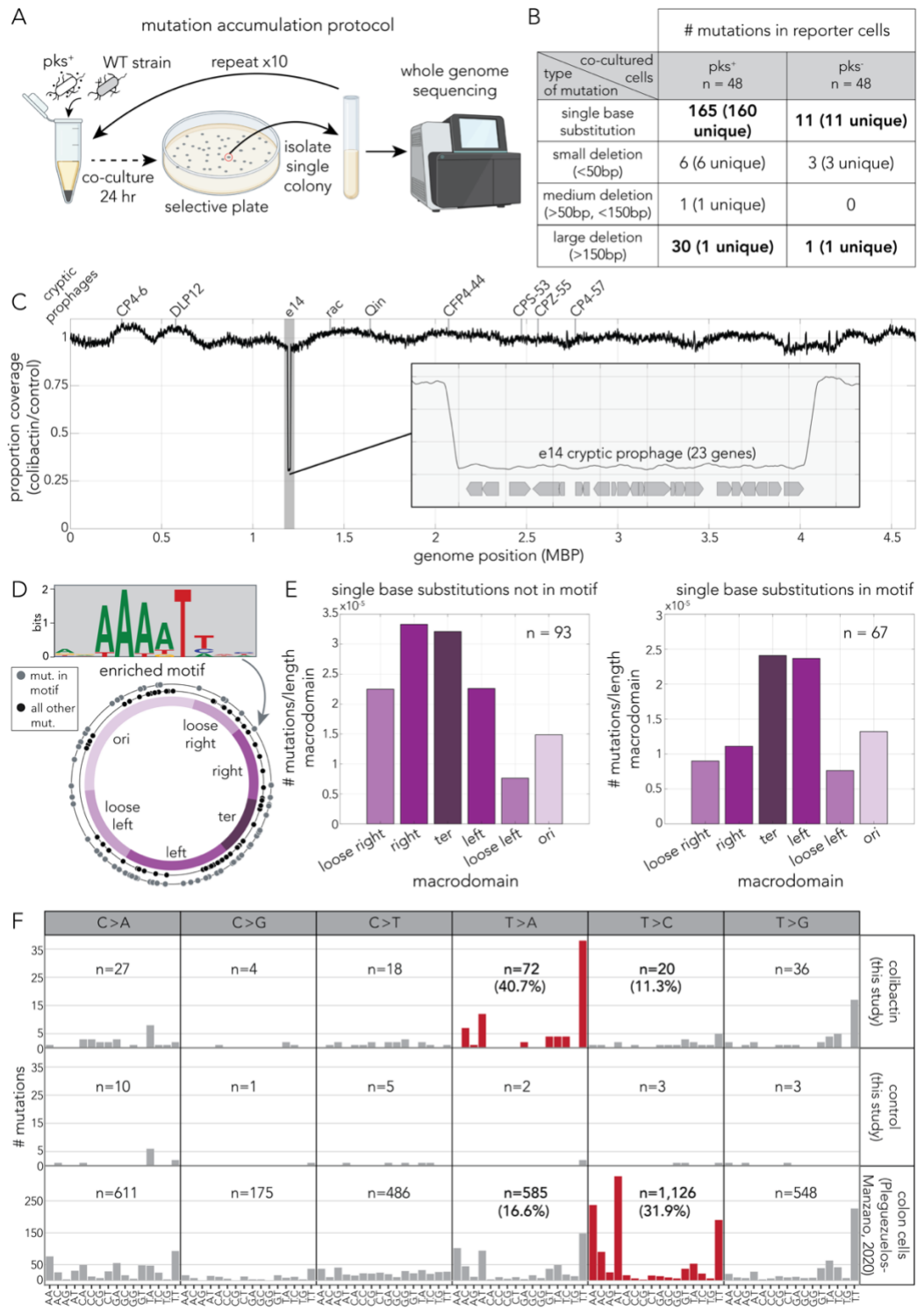


Figure 3.3 Colibactin induces a bacteria-specific mutational signature.

(A) Overview of mutation accumulation experiment. Reporter cells were co-cultured with pks⁺ cells in a pellet for 24 hours before plating on selective agar. Single reporter colonies were selected and grown for subsequent exposure.

Reporter cells were exposed ten times before whole-genome sequencing. (B) Summary table of annotated mutations by mutation type. (C) Whole-genome sequencing coverage of colibactin/control conditions. The positions for all nine cryptic prophages are marked above. The gray-shaded region marks the e14 prophage region shown in more detail in the panel inset. (D) The A/T enriched motif found in the 13 bp region that surrounds positions of single base substitutions. The ring plots show positions of single base substitutions (outer ring in gray – SBS matching the motif, inner ring in black – all other SBS). The chromosome ring shows previously defined macrodomain regions. (E) SBS positions show genomic positional bias and are enriched near the terminus, increasingly so for mutations occurring in the identified motif. (F) Trinucleotide context of SBS mutations. The upper panel shows mutations identified after co-culturing with *pks*⁺ bacteria. The middle panel shows mutations identified after co-culturing with *pks*⁻ bacteria (control). The bottom panel shows mutations annotated in colon cells exposed to *pks*⁺ bacteria by Pleguezuelos-Marzano et al. (2020). Differential mutation signatures between our work and work in colon cells are highlighted in red.

We analyzed the genomic context surrounding single base substitutions to check whether specific motifs are enriched for colibactin-induced mutations using the STREME (Bailey 2021) tool. Figure 3.3D shows the statistically significant motif (p-value = 0.003) that was found in 67 of 160 unique SBS. The A/T rich motif we found was similar to the motif identified in colorectal cancers (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). Within this motif, the mutated base was either the A in position 4 or the T in position 7. We next tested if enriched SBS positions show a spatial preference using previously annotated chromosome macrodomains (Lioy et al. 2018) shown in Figure 3.3D. We observed that all SBS sites, those matching the motif and those not matching it, were commonly closer to the terminus region (ter; Figure 3.3D,E). Finally, we directly compared the mutation patterns we found with those reported for colon organoids (Pleguezuelos-Manzano et al. 2020) by looking at the trinucleotide context of the mutated base (Figure 3.3F). This comparison showed that despite

similarity in the genomic context, a striking difference exists in the mutation outcome. As Figure 3F shows, mutations in A/T rich trinucleotides in *E. coli* were predominantly T>A while in colon cells they were predominantly T>C.

In summary, we found similarities, but also clear differences, between colibactin-induced mutations in bacteria and colon cells. In both cases, increased mutation rates are clearly detected and show a bias towards A/T rich sequence motifs. However, differences exist in the nucleotide that results from the mutation event and far reduced frequency of indels in *E. coli* relative to colon cells. Lastly, we detected that colibactin-induced mutations were biased in some genome macrodomains in *E. coli*.

3.4.4 Self-inflicted damage in colibactin-producing *E. coli*

Colibactin self-protection is attributed to two mechanisms: colibactin activation by the ClbP peptidase after it is exported to the periplasm (Velilla et al. 2023; Dubois et al. 2011; Brotherton and Balskus 2013) and inactivation of intracellular colibactin by the ClbS cyclopropane hydrolase (Bossuet-Greif et al. 2016; Tripathi et al. 2017). Given that we did not identify genes involved in colibactin import in our genetic screen, we hypothesized that mature colibactin may permeate back into producing cells. We therefore further hypothesized that producer cells might experience weak, yet elevated levels, of DNA damage despite self-protection mechanisms. We first tested this hypothesis by introducing a fluorescent DNA damage reporter into *pks*⁺ and *pks*⁻ strains. We constructed the

fluorescent DNA damage reporter by cloning a YFP fused to the *recA* gene promoter, known to respond to DNA damage (Vollmer et al. 1997), and a CFP under control of a constitutive promoter into a low-copy plasmid (Figure 3.4A). The CFP tag allowed us to identify reporter cell colonies and normalize DNA damage-induced fluorescence (YFP) to cell numbers in reporter colonies.

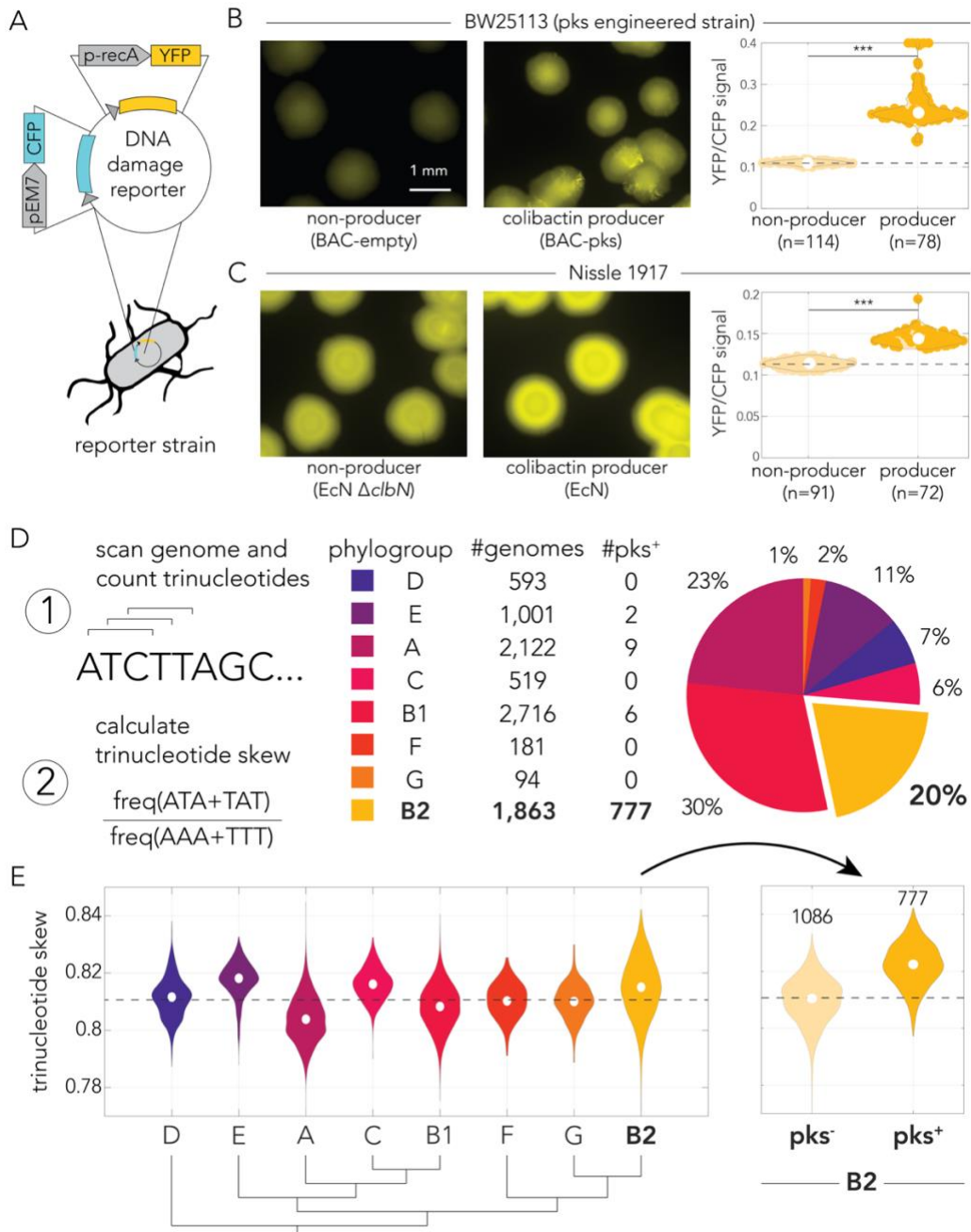


Figure 3.4 Colibactin inflicts self-damage.

(A) Schematic of the fluorescent DNA damage reporter plasmid used to quantify self-damage. CFP is expressed under a constitutive promoter while YFP is expressed under a DNA damage-inducible promoter. (B) Representative microscopy images of engineered (BAC) pks^+ and pks^- colonies or (C) Nissle 1917 pks^+ (EcN) and pks^- (EcN $\Delta clbN$) expressing our *recA* reporter plasmid. Images show YFP expression, which represents *recA* activation in the colonies due to colibactin-induced self-damage. Fluorescence intensity range was set according to the intensity observed in the colibactin-producer field of view per strains. Violin plots display the median YFP signal intensity per colony for each strain. Background YFP and CFP autofluorescence of the colonies was subtracted from each channel before YFP was normalized to CFP per colony (** $p < 0.001$, two sample *t*-test). (D) Schematic of our analysis of 9,089 genomes to calculate the trinucleotide skew towards ATA/TAT over AAA/TTT sequences. Genomes are broken down by phylogroup and the total number of genomes per phylogroup, as well as the number of pks^+ genomes, included in our analysis are reported. The pie chart shows the total percentage of genomes each phylogroup represented. (E) (left) Violin plots of trinucleotide skewness by *E. coli* phylogroup. Median skew is marked by the dashed line. (right) The B2 phylogroup skewness is further divided into genomes with colibactin and genomes without colibactin. The number of genomes in each condition is labeled in this panel.

These strains were then grown on agar to quantify the intensity of the fluorescent reporter. Figure 3.4B shows representative microscopy images of colibactin-producers and non-producers. In agreement with our hypothesis, we indeed observed increased DNA damage reporter activity in cloned cells harboring the *pks* pathogenicity island relative to control non-producers. Quantification of the median reporter level across dozens of colonies revealed this increase is statistically significant (Figure 3.4B, right panel). Since colibactin expression level in genetically engineered clones may be higher than expression levels of strains naturally harboring the *pks* pathogenicity island, we repeated this experiment with the Nissle 1917 strain that naturally expresses colibactin (Figure 3.4C). We again observed elevated reporter activity in colibactin-producers relative to non-producing *clbN* knockouts (Figure 3.4C, right panel). As expected,

this increase was relatively weak when compared to the engineered *pks*⁺ strain in Figure 3.4A as well as a *clbS* knockout that is unable to inactivate colibactin to protect the producing cells (Figure 3.5).

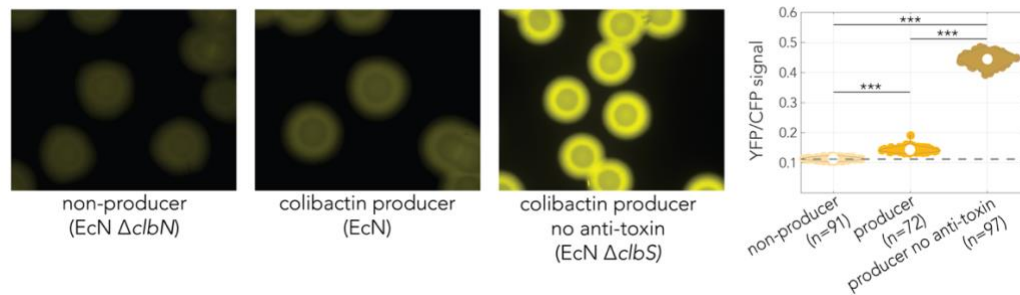


Figure 3.5 Colibactin inflicts more self-damage in the absence of ClbS.

Representative microscopy images of Nissle 1917 *pks*⁻ (EcN Δ *clbN*), *pks*⁺ (EcN), and *pks*⁺ (EcN Δ *clbS*) expressing our *recA* reporter plasmid. Images show YFP expression, which represents *recA* activation in the colonies due to colibactin-induced self-damage. Fluorescence intensity range was set according to the intensity observed in the most highly damaged colibactin-producer (EcN Δ *clbS*) field of view. Violin plot displays the median YFP signal intensity per colony for each strain. Background YFP and CFP autofluorescence of the colonies was subtracted from each channel before YFP was normalized to CFP per colony (***) $p < 0.001$, two sample t-test).

Motivated by the observation that colibactin-producers show increased DNA damage reporter activity, we hypothesized that self-inflicted damage may also be evident in the genome of strains harboring the *pks* pathogenicity island. We reasoned that this prediction can be evaluated by examining whether there is a bias in trinucleotide sequences in genomic DNA. Specifically, we predicted colibactin-induced DNA damage will be associated with a bias toward ATA and TAT trinucleotides at the expense of AAA and TTT trinucleotides (Figure 3.3F). In a random genome, we would expect equal proportions of these trinucleotides, thus shifts in these proportions can be measured as an indicator of colibactin

induced DNA damage. We validated this expectation in a mock random genome using a sliding window of three nucleotides to calculate the proportion of trinucleotides matching ATA/TAT or AAA/TTT. This simulation yielded equal proportions of each complementary trinucleotide sequence, suggesting that using a sliding window does not bias our results.

We tested our hypothesis in 9,089 annotated genomes spanning all *E. coli* phylogroups (Figure 3.4D) and calculated a trinucleotide skewness metric $(N_{ATA}+N_{TAT})/(N_{AAA}+N_{TTT})$ to detect any bias towards colibactin-associated trinucleotide sequences. This metric is not strand-specific as we pooled together complementary trinucleotides in this calculation (e.g., $N_{ATA}+N_{TAT}$). The strains included in this analysis and their calculated trinucleotide skew are reported in Supplementary File 7. Of the 9,089 genomes analyzed, 794 were annotated as containing the *pks* island, with the majority (777) belonging to the B2 phylogroup. The almost exclusive occurrence of *pks*⁺ strains in the B2 phylogroup agrees with previous findings (Nougayrède et al. 2006; Wami et al. 2021). Figure 3.4E shows the skewness levels calculated for all genomes classified by phylogroup.

Calculating the skewness for each phylogroup allowed for setting an expectation on the amount of skew in a phylogroup without the *pks* island. In agreement with our prediction, we found that the B2 phylogroup, as the most prevalent phylogroup containing the *pks* island (Nougayrède et al. 2006; Wami et al. 2021), had one of the largest ranges of skew with the trinucleotide skew metric.

Moreover, separating the strains of this phylogroup to *pks*⁺ and *pks*⁻ subgroups

(Figure 3.4D, right panel) revealed that the skewness is significantly increased specifically in *pks*⁺ strains (p-value < 10⁻¹⁰). The range of the skewness in the separated *pks*⁺ strains and *pks*⁻ strains in the B2 phylogroup is similar to the other *E. coli* phylogroups. Taken together, our microscopy and genomic analysis results indicate that colibactin-producing strains likely experience elevated levels of basal DNA damage culminating in noticeable bias in trinucleotide genomic composition that is compatible with colibactin self-inflicted damage.

3.5 Discussion

Colibactin-producing bacteria are not uncommon in the gut microbiome of healthy humans, yet their increased prevalence is evident in multiple human diseases ranging from Inflammatory bowel disease to colon cancers (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). Compelling evidence from colorectal tumors in humans strongly supports the premise that colibactin acts as a tumorigenic mutagen (Dziubańska-Kusibab et al. 2020; Pleguezuelos-Manzano et al. 2020). The clinical relevance of colibactin-induced damage has motivated intense research into the mechanisms underlying its toxicity in eukaryotes, but left fundamental questions underexplored in bacteria. Our study addressed some of these gaps in knowledge and revealed differences between the bacterial and mammalian cellular response to colibactin.

Our study of the cellular mechanisms underlying colibactin damage relied on a loss-of-function genetic screen. This genome-wide approach, applied for the first time to study colibactin toxicity, was unbiased by current understandings of

colibactin mode of action. We identified various DNA damage response pathways whose inactivation increased colibactin sensitivity, uncovering a crucial role for homologous recombination (Figure 3.2A,C). We only detected a single hit from the nucleotide excision repair pathway, despite its known role in removing interstrand crosslinks (Cole 1973; Bossuet-Greif et al. 2016). The findings from our genome-wide genetic screen strongly agree with the current understanding that colibactin-induced toxicity is due to its role as a DNA-damaging agent. This DNA damage likely underlies the frequent excision of prophages that we (Figure 3.3C) and others observed (Silpe et al. 2022).

A novel finding of our work emerged from mutation accumulation experiments. We found that in bacteria, similar to eukaryotes (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020), single base substitution mutations were primarily located in A/T rich DNA sequences (Figures 3.4D and 3.4F). This similarity can be rationalized by the structural model suggesting that colibactin targets these A/T rich sequences due to their particularly narrow minor groove (Dziubańska-Kusibab et al. 2020). Interestingly, despite a similar binding preference, colibactin-induced mutations were markedly different: in colon cells T>C was the predominant change, while T>A was most prevalent in bacteria (Figure 3.4F). The difference in DNA repair mechanisms between bacteria and mammalian cells is one plausible mechanism that may underly this mutational dissimilarity. Additionally, indel mutations are abundant in colon cells

(Pleguezuelos-Manzano et al. 2020) but uncommon in bacteria. We also detected a positional bias in colibactin-induced mutations in bacteria (near the terminus).

Our genetic screen uncovered that homologous recombination, an error-free repair pathway, plays a major role in mitigating colibactin-induced DNA damage.

However, results from our mutation accumulation experiments revealed that colibactin is mutagenic in *E. coli*. This seeming inconsistency may be explained by the complexity of the DNA damage response colibactin induces in *E. coli* that combines both error-prone and error-free repair mechanisms. Following activation of the SOS response, nucleotide excision repair, an error-free mechanism, repairs damage, but extensive damage can activate polymerase V involved in translesion synthesis, which is error-prone (Maslowska et al. 2019).

Our screen results suggested that all these pathways (homologous recombination, nucleotide excision repair, and translesion synthesis) indeed participate in mitigating colibactin induced damage. It is also intriguing to note that colibactin induces a specific SBS profile. Different SBS patterns were previously reported for other DNA damaging agents such as UV and mitomycin-C (Maslowska et al. 2019; Kowalczykowski et al. 1994). The divergent mutation patterns reported for these different agents likely arises from a combination of diverse responses to the types of damage and the different DNA sequences being impacted by each of the agents.

Finally, we leveraged the bacteria-specific mutational bias we discovered to investigate if colibactin inflicts self-damage in producers. We found that in both engineered and naturally producing cells, colibactin inflicts self-toxicity that was visualized and quantified with a fluorescent DNA damage reporter (Figure 4B, 4C). This is likely due to small quantities of the toxin re-entering the producing cells and binding DNA before it is inactivated by the ClbS cyclopropane hydrolase. Excitingly, comparative analysis of almost 2,000 *E. coli* genomes from the B2 phylogroup provides supporting evidence of colibactin-linked skewness in trinucleotide composition in strains harboring the *pks* pathogenicity island. Reassuringly, the skewness in *pks*⁺ strains of the B2 phylogroup was highest among all phylogroups of the *E. coli* species. Hence it seems that colibactin entails a cost on producing cells and leaves an evolutionary footprint in their genomes.

Taken together, our systematic work reveals unique features of colibactin toxicity in bacteria. The focus of this study on the effects of colibactin in bacteria is important given that this toxin is commonly found in non-pathogenic strains and therefore likely emerged, like many other bacterial toxins, to facilitate competition within microbial communities (Kern et al. 2021). Our study outlines an important direction for future investigation. This key direction builds on the newly identified bacteria-specific mutational signature. Specifically, it will be interesting to explore if evidence of colibactin-specific signatures can be found in longitudinal microbiome samples from individuals that harbor colibactin-

producing bacteria in their gut. Quantifying the strength and rate of a colibactin-linked mutational bias may provide a noninvasive method to estimate the intensity of colibactin exposure in a specific individual. Given that 30% of healthy individuals harbor *pks*⁺ strains in their gut microbiome, additional information about the rate of colibactin damage accumulation in the individual's microbiome may help to gauge their risk for developing colibactin-linked colon cancer.

3.6 Methods

Table 3.1 Bacteria strains used in this study

Strain	Nickname	Source	Use
BW25113 Δ <i>gspI::carb</i>	Viability reporter strain; ampicillin-resistant reporter strain	This study	Colibactin impact on cell viability, mutation accumulation
BW25113 pBeloBAC11+pks	Engineered <i>pks</i> ⁺	Silpe et al. 2022 (Silpe et al. 2022)	Colibactin impact on cell viability, genetic screens
BW25113 pBeloBAC11	Engineered <i>pks</i> ⁻	Silpe et al. 2022 (Silpe et al. 2022)	Colibactin impact on cell viability, genetic screens
Pooled <i>E. coli</i> library	Pooled <i>E. coli</i> library	Hirotsuda Mori, Nara Institute of Science and Technology, Japan	Genetic screens
BW25113 pBeloBAC11+pks pRecA	Engineered <i>pks</i> ⁺ with DNA damage reporter	This study	Self-inflicted damage
BW25113 pBeloBAC11 pRecA	Engineered <i>pks</i> ⁻ with DNA damage reporter	This study	Self-inflicted damage
Nissle 1917 pRecA	Nissle 1917 <i>pks</i> ⁺ DNA damage reporter	This study	Self-inflicted damage
Nissle 1917 Δ <i>clbN::chl</i> pRecA	Nissle 1917 <i>pks</i> ⁻ DNA damage reporter	This study	Self-inflicted damage

Nissle 1917 $\Delta clbS::chl$ pRecA	Nissle 1917 pks ⁺ <i>clbS</i> knockout DNA damage reporter	This study	Self-inflicted damage
---	--	------------	--------------------------

3.6.1 Media and growth conditions

All experiments were performed in either Luria Broth (LB) or minimal synthetic media (M9 salts supplemented with 0.4% glucose, 2 mM MgSO₄, 0.1 mM CaCl₂, 0.2% ampicase). Overnight cultures for all experiments were grown at 37°C with 200 rpm orbital shaking. During the overnight growth of antibiotic-resistant strains, we added antibiotics at the following concentrations: 50 µg/mL spectinomycin, 50 µg/mL kanamycin, 25 µg/mL chloramphenicol, and 50 µg/mL carbenicillin.

3.6.2 Cloning deletion strains

gspI was deleted from BW25113 using lambda red recombination with an insert containing 60 bp homology arms to the upstream and downstream genomic regions and a carbenicillin resistance cassette. *clbN* and *clbS* were deleted from Nissle 1917 using lambda red recombination with an insert containing 40 or 60 bp homology arms to the upstream and downstream genomic regions, respectively, and a chloramphenicol resistance cassette.

3.6.3 Monitoring colibactin impact on viability of co-cultured cells

Cultures of the viability reporter strain and the engineered pks⁺ and pks⁻ strains were grown overnight in LB. 1 mL of overnight culture was washed three times

with PBS. Cultures were then diluted 1:50 into either LB or M9 and grown for two hours. Following growth, OD₆₀₀ was measured, and cultures were diluted to OD₆₀₀=0.1. Reporters and engineered pks⁺ and pks⁻ strains were mixed in 96 deep-well plates (Eppendorf, cat# 2231000920) to a final volume of 500 μL at a 10:1 or 1:1 ratio (producers to reporters). For the pelleted co-culture conditions, plates were centrifuged at 4000 g for 6 minutes and incubated at 37°C with no shaking. For suspension co-culture conditions, plates were incubated at 37°C with 200 rpm orbital shaking.

At each time point, co-cultures were thoroughly mixed by pipetting them and 5 μL aliquots were transferred to the top row of a 96-well microplate (FisherBrand, cat# FB012932) containing 95 μL PBS. Samples were serially diluted 1:5 over the seven remaining rows in the microplate. From each dilution well, 4 μL was spotted on LB agar plates supplemented with spectinomycin to select for the reporter strain. We then back-calculated CFUs in each sample based on the dilution factor of the least dilute spot for each sample that contained 3-25 colonies.

3.6.4 Genome-wide loss-of-function genetic screen

We used a pooled genetic screening approach that we previously developed (Rosener et al. 2020; Sayin et al. 2023; Noto Guillen et al. 2021, 2024) to identify genes and pathways impacting colibactin sensitivity. The method relies on a collection of 7,259 knockout strains that span 3,680 non-essential genes in *E. coli*.

Each knockout strain harbors a 20 bp nucleotide barcode that is integrated into its chromosome. This collection allows for performing pooled genetic screens and identifying the frequency of each knockout strain by targeted deep sequencing of the barcode locus.

A 200 μ L aliquot of frozen glycerol stocks of the pooled *E. coli* library was grown overnight in LB supplemented with chloramphenicol. The engineered pks⁺ and pks⁻ strains were also grown overnight in LB supplemented with chloramphenicol. Cultures were washed twice in PBS and resuspended in M9. The knockout collection was diluted 1:50 in 40 mL M9. The engineered pks⁺ and pks⁻ strains were diluted 1:50 in 100 mL M9. Cultures were grown for two additional hours at 37°C and 200 rpm to allow for adjustment to media and exit from stationary phase before adjusting their density to OD₆₀₀=0.1 in 60 mL for the knockout collection and 140 mL for the engineered strains. The cultures were mixed at a 1:1 and 10:1 ratio and then divided into 8 mL replicates spread across 16 wells in a 96 deep-well plate (Eppendorf, cat# 2231000920). Co-cultures were pelleted at 4000 g for 6 minutes before being incubated at 37°C for 8, 24 and 48 hours. Cultures were then resuspended and all wells per replicate were merged into a conical 50 mL tube. The cultures were pelleted at 4000 g for 6 minutes and media was aspirated. Pellets were flash frozen and stored at -80°C until DNA was extracted using Zymo Quick-DNA Midiprep Plus Kit (cat# D4075).

DNA concentrations were measured with Quant-iT dsDNA high sensitivity assay (Invitrogen, cat# Q33232) on a Tecan Spark plate reader and samples were normalized to 20 ng/ μ L. The barcode sequencing protocols we previously developed were modified to account for the low proportion of DNA in each sample that originated from the knockout library: DNA template for each PCR reaction was increased to 20 ng and 4 reactions were set up per sample to get sufficient coverage of the library (based on calculations of the final CFUs of reporter cells and pks⁺ cells in spotting co-cultures). The template DNA was amplified over 23 cycles with custom forward and reverse primers and 2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, cat# KK2602). Following barcode amplification, the 4 reactions were pooled, and 25 μ L was purified for downstream use. The PCR product was purified with AMPure XP beads (Beckman Coulter, Cat#A63881) following the standard protocol of beads added at 0.9x sample volume. We modified the Nextera XT Index Kit (Illumina, cat# FC-131-1024) protocol to work with half volumes. These products were then run on a 2.5% agarose gel and extracted using ZR-96 ZymoClean Gel Recovery Kit (cat# D4021). The purified libraries were quantified with the Quant-iT dsDNA high sensitivity assay and normalized to 4 nM. Library quality was assessed on a Bioanalyzer with the Agilent High Sensitivity DNA Kit (Agilent Technologies, Cat# 5067-4626). Libraries were denatured and diluted according to the NextSeq 500/550 system protocol and sequenced with the NextSeq 500/550 High Output kit v2.5, 75 cycles (Illumina, cat# 20024906) on a NextSeq 500/550 machine.

3.6.5 Analysis of genetic screen results

We extracted barcode counts from FASTQ sequencing files using a custom MATLAB (MathWorks) script. Exact matches of barcodes (15-25 bp) were searched for in each read and a knockout strain was assigned if there was a matching barcode. Any nucleotide with a quality score below 10 was masked in the analysis. Counts of knockout strains targeting the same gene were summed together. Knockouts that impacted sensitivity to colibactin were identified by comparing the relative frequency of a knockout strain in the pks⁺ co-culture condition with the relative frequency in the pks⁻ co-culture condition. The statistical significance of the changes in relative frequency was determined with the DESeq2 tool (Love et al. 2014). We chose log₂-fold-change (>1.3) and adjusted p-value thresholds (<0.25) to classify resistant and sensitive knockouts. We used the gene set enrichment analysis tool GAGE (Luo et al. 2009) to test for functional enrichment. For this analysis we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) and Gene Ontology (GO) (The Gene Ontology Consortium et al. 2000) databases.

3.6.6 Targeted validation genetic screen

Knockout strains for all resistant and sensitive hits determined by the genome-wide genetic screen (including several marginally sensitive (4) or resistant (5) strains) along with 38 neutral strains were picked from glycerol stocks of single knockout strains composing the genome-wide library. Strains were cultured in 1 mL LB supplemented with chloramphenicol overnight. The engineered pks⁺ and

pk^s strains were also grown overnight in LB supplemented with chloramphenicol. The following day, 100 μ L of each resistant or sensitive knockout strain culture was combined along with 400 μ L of each neutral knockout strain to pool all strains together to a single culture. From the combined strains, 2 mL was taken to wash, along with 2 mL of each engineered overnight culture. Cultures were washed twice in PBS before resuspending in M9 media. Both engineered strains were diluted 1:50 in 100 mL of M9. The pooled knockout strains were diluted 1:50 in 45 mL of M9. The diluted cultures were incubated and grown for three hours to adjust to the M9 media and exit stationary phase. Then the culture density was measured with OD₆₀₀ and cultures were diluted to OD₆₀₀=0.1. The cultures were mixed in 10:1 (producer to reporter) ratios and divided into 8 mL replicates spread across 16 wells in a 96 deep-well plate (Eppendorf, cat# 2231000920). Co-cultures were pelleted at 4000 g for 6 minutes before being incubating at 37°C for 48 hours (after 24 hours cultures were mixed and re-pelleted). Cultures were then resuspended and all wells per replicate were merged into a conical 50 mL tube. The cultures were pelleted at 4000 g for 6 minutes and media was aspirated. Pellets were flash frozen and stored at -80°C until DNA was extracted using Zymo Quick-DNA Midiprep Plus Kit (cat# D4075). DNA Library preparation and targeted sequencing was identical to the genome-wide screen.

3.6.7 Mutation accumulation

An ampicillin-resistant BW25113 strain and the engineered pks^+ and pks^- strains were grown overnight on LB supplemented with antibiotics. 1 mL aliquots of each overnight culture were washed in PBS three times before 1:50 dilution in M9 followed by growth for two hours at 37°C. Cultures were then normalized to $\text{OD}_{600}=0.1$. Reporters and engineered pks^+ and pks^- strains were mixed in 96 deep-well plates (Eppendorf, cat# 2231000920) to a final volume of 500 μL at a 10:1 ratio (producers to reporters) and pelleted by centrifuge at 4000 g for 6 minutes. The cultures were incubated at 37°C without shaking for 24 hours. Wells were then mixed and 5 μL of each co-culture was transferred to a 96 well microplate (FisherBrand, cat# FB012932) and diluted by mixing with 95 μL of PBS. From there, each culture was serially diluted 1:5 for 7 dilutions total. 4 μL from each dilution was spotted on LB agar plates supplemented with carbenicillin to select for the ampicillin-resistant BW25113 strain. The following day, a single colony was picked from each replicate from the spotting agar plate and inoculated into 500 μL M9 in a 96 deep-well plate. Fresh cultures of the engineered pks^+ and pks^- strains were grown overnight and the co-culture protocol was repeated. This process, from diluting the 24-hour co-cultures, spotting, expanding single colonies, and setting up new co-cultures, was repeated for 10 cycles. After that, single colonies were picked from each replicate, expanded, and frozen as glycerol stocks.

Glycerol stocks were inoculated into 1 mL LB and grown for 3-4 hours before DNA was extracted with the Zymo Quick DNA 96 kit (cat# D3012) and quantified with Quant-iT dsDNA high sensitivity assay (Invitrogen, cat# Q33232). Samples were normalized to 12 ng/ μ L. Prior to sequencing, four replicates from each co-culture condition were pooled evenly by combining 10 μ L of each 12 ng/ μ L stock. The pooled samples (24 total) were sent for whole-genome sequencing at SeqCenter (Illumina paired-end sequencing with 2x151bp). The average coverage per genome was 65. Reads were aligned to the reference genome (NCBI accession: CP009273) with breseq (Deatherage and Barrick 2014) to identify mutations. The tool was used in population mode (we discarded mutations with a proportion below 0.1). Mutations occurring in both the pks^+ and pks^- conditions were excluded as they likely existed in the ancestor. To annotate large deletions, we determined genome coverage of each sample with a custom Python script using a sliding window of 10,000 bp. We examined the coverage by plotting it after further smoothing in MATLAB (MathWorks). Sequence encompassing 6 bp upstream and downstream of single-base substitutions was analyzed for enriched motifs using STREME(Bailey 2021). Since only a few mutations were identified in the pks^- condition, we selected 1000 random 13 bp sequences from the reference genome to use as a control. Sequences containing the enriched motif were identified using FIMO(Grant et al. 2011) with the probability matrices generated by STREME. The frequency of the enriched motif sequences was plotted across previously defined macrodomains in the *E. coli* genome (Lioy et al. 2018).

3.6.8 DNA damage fluorescent reporter

We cloned the DNA damage reporter plasmid with the Gibson assembly method (Gibson et al. 2009) using In-Fusion Snap Assembly Master Mix (Takara, cat# 638947). In a single assembly reaction, we integrated a YFP and the *recA* promoter into a plasmid backbone containing a spectinomycin resistance cassette and CFP. When amplifying the backbone, we also replaced the CFP promoter with a 48 bp EM7 promoter that was encoded on the amplification primer. The 81 bp *recA* promoter was amplified from the BW25113 genome (the promoter region was defined according to previous work (Salgado et al. 2023; Pagès et al. 2003; Stohl et al. 2003)). The final plasmid was a low-copy plasmid with the *sc101* origin and spectinomycin resistance. Plasmid assembly was validated with Sanger sequencing spanning the integration sites.

3.6.9 Self-inflicted damage microscopy

The engineered *pks*⁺ and *pks*⁻ strains and the Nissle strains (*pks*⁺, *pks*⁻, and *pks*⁺ *clbS* knockout) harboring the DNA damage reporter were grown overnight in LB supplemented with antibiotics. OD₆₀₀ was measured and each culture was subsequently diluted so that plating on agar plates would yield ~200 colonies. The diluted cultures were spread with glass beads on M9 agar plates and were incubated at 37°C. We imaged 40-70 fields of view for each plate with a Zeiss Axio Observer.Z1 epifluorescence microscope. Colonies were imaged at 2.5x magnification using CFP (475nm), YFP (524nm), and brightfield channels.

Analysis was performed using custom MATLAB scripts (MathWorks). Briefly, colonies were segmented using an automatically determined threshold on the CFP channel image. To ensure that we only analyzed whole, single colonies, we excluded masked regions contacting the image border and filtered masks by area and circularity. Within each mask, the median YFP signal was measured. Background autofluorescence of untagged colonies was subtracted from both YFP and CFP before the YFP signal of each colony was normalized to its CFP signal.

3.6.10 Bioinformatics analysis of colibactin-linked trinucleotide skew

We tested the trinucleotide composition of 9,089 *E. coli* genomes downloaded from the NCBI genome database that were previously assigned to a specific phylogroup (Abram et al. 2021). We discarded genome assemblies that had genome lengths below 4 Mb and those that were flagged as “Status: suppressed” by RefSeq annotation (this flag points to potential concerns with the genome assembly). For each genome, we scanned the DNA sequences of all contigs and calculated the frequencies of all 64 possible trinucleotides. For each genome we computed the colibactin-linked skewness metric by calculating the ratio of sum frequencies of the complementary trinucleotides ATA and TAT relative to the sum frequencies of the complementary trinucleotides AAA and TTT ($\text{Skew} = (\text{N}_{\text{ATA}} + \text{N}_{\text{TAT}}) / (\text{N}_{\text{AAA}} + \text{N}_{\text{TTT}})$). It is important to note that the nucleotide compositions on the trinucleotides in the numerator and denominator is equal (three A’s and three T’s). Therefore, our skewness measurement reflects bias in nucleotide order and is indifferent to differences in nucleotide composition. We

classified strains as colibactin producers by the proteome annotation associated with each genome assembly. Strains harboring more than nine proteins annotated with “colibactin” in their description were classified as pks⁺ strains. Our choice of at least nine pks genes (~50% of the pks genes) was used to account for potential missing coverage of the pks genes in the deposited genome sequences. Of the 19 genes in the island, 17 are considered essential for colibactin expression (Nougayrède et al. 2006). Out of the deposited genomes we analyzed, 794 had more than nine pks genes and 98% of these (777) had 17 or more pks genes. Out of a total of 9,089 *E. coli* genomes we analyzed, 794 strains were classified as pks⁺ with 98% (n=777) of them belonging to the B2 phylogroup. The remaining pks⁺ strains belonged to the E (n=2), A (n=9), and B1 (n=6) phylogroups. The B2 phylogroup is known to harbor the vast majority of pks⁺ strains in *E. coli* (Nougayrède et al. 2006; Wami et al. 2021).

3.6.11 Data Availability

All raw sequencing data from the barcoded knockout library screen and whole genomes from the mutation accumulation experiment are available in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession numbers PRJNA1061230, PRJNA1060772, and PRJNA1060778. Code for all analyses can be found at GitHub (https://github.com/Mitchell-SysBio/2023_colibactin) (DOI: 10.5281/zenodo.10419723) and in the Supplemental Materials.

Chapter 4: Colibactin-induced damage in bacteria is cell contact independent

4.1 Preface

The work in this chapter has been adapted from the following publication:

Lowry E, Mitchell A. 2024. Colibactin-induced damage in bacteria is cell contact independent. *bioRxiv* 2024.06.21.600066.

<http://biorxiv.org/content/early/2024/06/21/2024.06.21.600066.abstract>.

The author contributions are as follows:

E.L.—Conceptualization, Experiments, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing and Editing;

A.M.—Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing and Editing.

4.2 Abstract

The bacterial toxin colibactin, produced primarily by the B2 phylogroup of *Escherichia coli*, underlies some cases of colorectal cancers. Colibactin crosslinks DNA and induces genotoxic damage in both mammalian and bacterial cells.

While the mechanisms facilitating colibactin delivery remain unclear, results from multiple studies supported a delivery model that necessitates cell-cell contact. We directly tested this requirement in bacterial cultures by monitoring the spatiotemporal dynamics of the DNA damage response using a fluorescent transcriptional reporter. We found that in mixed-cell populations, DNA damage

saturated within twelve hours and was detectable even in reporter cells separated from colibactin producers by hundreds of microns. Experiments with distinctly separated producer and reporter colonies revealed that the intensity of DNA damage decays similarly with distance regardless of colony contact. Our work reveals that cell contact is inconsequential for colibactin delivery in bacteria and suggests that contact-dependence needs to be reexamined in mammalian cells as well.

4.3 Importance

Colibactin is a bacteria-produced toxin that binds and damages DNA. It has been widely studied in mammalian cells due to its potential role in tumorigenesis. However, fundamental questions about its impact in bacteria remain underexplored. We used *E. coli* as a model system to study colibactin toxicity in neighboring bacteria and directly tested if cell-cell contact is required for toxicity, as has previously been proposed. We found that colibactin can induce DNA damage in bacteria hundreds of microns away and that the intensity of DNA damage presents similarly regardless of cell-cell contact. Our work further suggests that the requirement for cell-cell contact for colibactin-induced toxicity also needs to be reevaluated in mammalian cells.

4.4 Introduction

Competitive interactions are prevalent within microbial communities, such as those found in the human gut microbiome (Kern et al. 2021). A common

mechanism facilitating these interactions involves the secretion of toxins that target neighboring microbes. One such toxin is colibactin, which is produced by certain bacteria and can bind to the DNA of nearby cells (Vizcaino and Crawford 2015; Pleguezuelos-Manzano et al. 2020; Nougayrède et al. 2006). Colibactin can induce damage to various bacterial species (Silpe et al. 2022; Wong et al. 2022; Chen et al. 2022) and can also affect host intestinal cells (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020). This genotoxin has been linked to numerous human diseases, including inflammatory bowel disease and colorectal cancers (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). While colibactin toxicity in host cells has been thoroughly investigated due to its clinical relevance, its ability to damage bacteria remains underexplored. Investigating colibactin toxicity in bacteria could elucidate how colibactin impacts the microbiota while also damaging host cells. Here, we study spatial and temporal dynamics of colibactin toxicity using the *Escherichia coli* model. We discovered that colibactin does not require cell-cell contact for toxicity, as previously suggested.

Colibactin is encoded by a 54kb genomic region known as the *pks* island. This island contains 19 genes needed to synthesize and export the toxin, including non-ribosomal peptide synthetases and polyketide synthases (Nougayrède et al. 2006). The island also encodes a cyclopropane hydrolase (*clbS*) that protects colibactin producers from self-inflicted damage (Nougayrède et al. 2006; Bossuet-Greif et al. 2016; Tripathi et al. 2017). The toxin itself contains two cyclopropane rings

that alkylate DNA and cause interstrand crosslinks (Xue et al. 2019; Tripathi et al. 2017; Xue et al. 2018; Vizcaino and Crawford 2015; Bossuet-Greif et al. 2018; Wilson et al. 2019). ClbS cleaves the cyclopropane rings to deactivate colibactin. The *pks* pathogenicity island is most commonly found in *E. coli* strains belonging to the B2 phylogenetic group and is expressed by both pathogenic and commensal strains (Nougayrède et al. 2006; Wami et al. 2021).

Colibactin toxicity has been broadly studied in mammalian cells due to the clinical relevance of the toxin and its association with various human diseases (Dougherty et al. 2023; Chen et al. 2023; Dubinsky et al. 2020; Iyadorai et al. 2020; Watanabe et al. 2020; Vizcaino and Crawford 2015; Cuevas-Ramos et al. 2010; Reuter et al. 2018; Secher et al. 2013; Lopès et al. 2020; Lucas et al. 2020; Iftexhar et al. 2021; Wilson et al. 2019; Dziubańska-Kusibab et al. 2020; Thakur et al. 2019; Bossuet-Greif et al. 2018; Cougnoux et al. 2014a). Colibactin toxicity has also been observed in bacteria. Colibactin-producing bacteria lacking the ClbS anti-toxin demonstrated auto-toxicity, which became more pronounced in cells with a deletion in the nucleotide excision repair pathway (Tripathi et al. 2017; Bossuet-Greif et al. 2016). Colibactin has been shown to target other bacterial species, including several *Staphylococcus* species (Wong et al. 2022; Chen et al. 2022; Silpe et al. 2022), several *Vibrio* species, *Clostridium difficile*, and *Enterobacter aerogenes* (Chen et al. 2022). One recent study suggested that colibactin toxicity in multiple bacteria species is attributed to prophage excision (Silpe et al. 2022). However, prophage-cured *S. aureus* remain susceptible to

colibactin, indicating that toxicity mechanisms beyond prophage induction exist (Wong et al. 2022).

A key gap in knowledge in the field of colibactin-induced damage concerns its delivery route to target cells. Colibactin instability has further confounded the study of this toxin in the extracellular environment (Wernke et al. 2020; Xue et al. 2019; Wernke et al. 2021; Vizcaino and Crawford 2015). The mechanism of colibactin secretion from producer cells remains largely unknown, but some studies in mammalian cells revealed its presence in outer membrane vesicles (Cañas et al. 2016; Fábrega et al. 2017). Multiple studies in both bacteria and mammalian cells claimed cell-cell contact is critical for potent colibactin-induced toxicity (Bossuet-Greif et al. 2018; Nougayrède et al. 2006; Reuter et al. 2018; Silpe et al. 2022; Chen et al. 2022). Here, we directly tested this requirement by monitoring the spatiotemporal dynamics of DNA damage response in recipient bacteria cells. Using live-cell fluorescent reporters of the DNA damage response, we show that colibactin-induced damage occurs within a few hours of exposure. Using two different experimental co-culturing setups, and both engineered and naturally colibactin producing strains, we discover that colibactin-induced DNA damage is detectable hundreds of microns away from the producing bacteria. Our work reveals that cell contact is inconsequential for colibactin delivery in bacteria and suggests that contact dependence needs to be reexamined in mammalian cells as well.

4.5 Results

4.5.1 Colibactin induces DNA damage in *E. coli*

To evaluate DNA damage in bacteria neighboring colibactin producers, we constructed a transcriptional reporter in an *E. coli* lab strain (Figure 4.1A). We cloned into a low-copy plasmid a YFP fused to the *recA* gene promoter, known to respond to DNA damage (Vollmer et al. 1997). Response of our reporter to DNA damage was validated with mitomycin-C, a known DNA damaging agent that primarily causes interstrand crosslinks (Wang et al. 2010) (Figure 4.2). We cloned into the same plasmid a CFP expressed from an EM7 constitutive promoter. The colibactin-producing strain was obtained by transforming the same lab strain with a bacterial artificial chromosome (BAC) expressing the *pks* pathogenicity island (Silpe et al. 2022) (Figure 4.1A). The same strain, carrying an empty BAC, was used as control. We refer to these strains as *pks*⁺ and *pks*⁻ hereafter. Both strains were transformed with a high-copy plasmid expressing an mCherry fluorescent protein from a constitutive promoter (UV5). Expression of CFP and mCherry from constitutive promoters allowed us to identify producers and target cells within a co-culture population, while expression of the YFP allowed us to quantify the level of DNA damage in target cells.

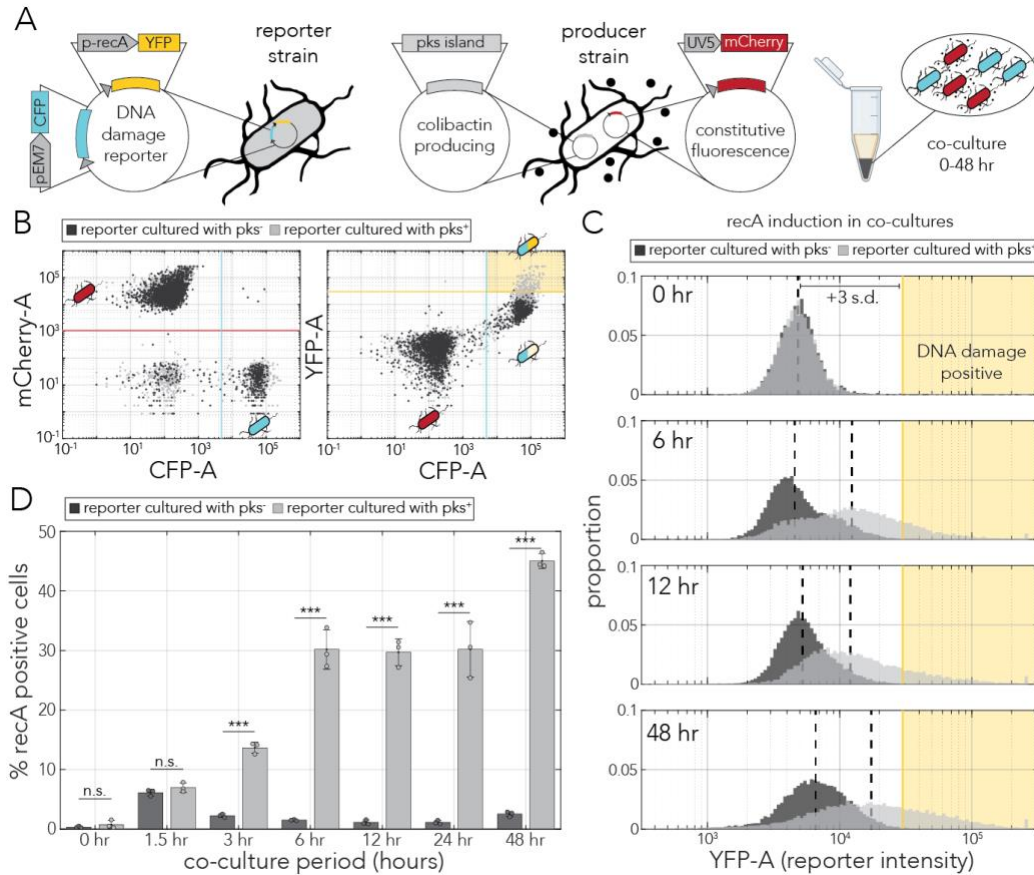


Figure 4.1 Temporal dynamics of colibactin-induced DNA damage.

(A) Plasmid design for tagging reporter cells and monitoring DNA damage response (left) and for tagging colibactin producers (right). Cells expressing each plasmid were co-cultured in a pellet for various time periods. (B) Flow cytometry gating strategy shown with representative cell populations (co-culture for 24 hr). A co-culture of reporters and producers ($p\text{ks}^+$) is marked in light gray and a control co-culture (reporters cultured with non-producers, $p\text{ks}^-$) is marked in dark gray. The left panel shows classification of cells as either reporters (high CFP) or producers (high mCherry). The right panel shows the YFP intensity, and the cutoffs used to identify the proportion of DNA damage positive reporters (yellow shaded rectangle corresponding to high CFP and high YFP signals). (C) Histograms of YFP intensity in reporter cells co-cultured with $p\text{ks}^+$ (light gray) and with $p\text{ks}^-$ (dark gray) over time. The yellow shaded area marks cells positive for a DNA damage response (three standard deviations above the mean signal at t_0). (D) Percent of DNA damage positive reporter cells when co-cultured with $p\text{ks}^+$ cells (light gray) and $p\text{ks}^-$ cells (dark gray). The error bars mark the standard deviations in three biological replicates and points represent the mean percent positive of each independent replicate (***) ($p < 0.001$).

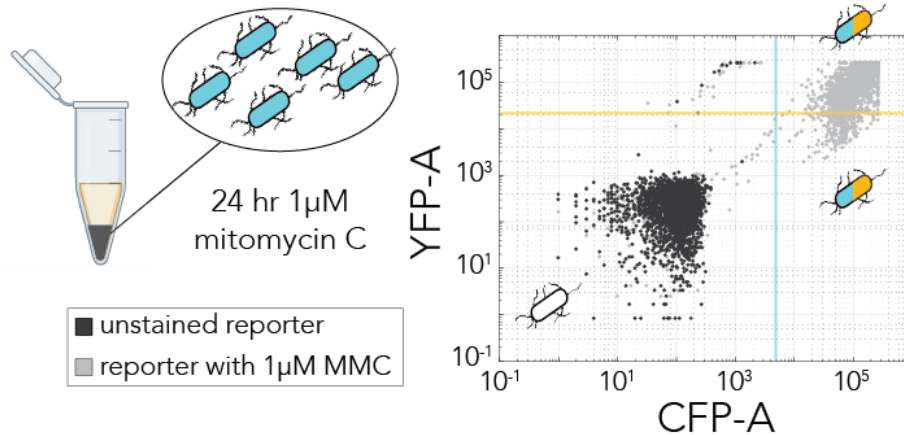


Figure 4.2 Validation of the *recA* DNA damage reporter.

Bacteria expressing the *recA* DNA damage reporter plasmid were incubated with 1 μM mitomycin-C (MMC) for 24 hours pelleted in minimal media. *E. coli* cells lacking the fluorescent plasmid served as an unstained control. Both cell populations were analyzed by flow cytometry. The same flow cytometry gating from Figure 1B is displayed on the scatter plot with the unstained population in dark gray and the MMC treated population in light gray.

We wanted to first test our reporter system and establish the timeline for DNA damage from colibactin in a co-culture model that maximized toxicity prior to testing the requirement for contact. We co-cultured colibactin-producing cells with reporter cells in a pellet at a 10:1 ratio for 48 hours (Figure 4.1A). We then used flow cytometry to identify populations of producer and reporter cells (Figure 4.1B, left panel). Using the separated populations, we then quantified the magnitude of the DNA damage response exclusively in reporter cells (Figure 4.1B, right panel). We chose a conservative threshold, three standard deviations above the YFP baseline, to determine the percentage of DNA damage positive cells (Figure 4.1B, yellow line).

Using the same threshold for all time points, we were able to evaluate the level of DNA damage in reporter cells over time. Figure 4.1C shows histograms of reporter fluorescence at selected time points. We observed a clear shift in the intensity histogram in pks^+ co-cultures, in contrast to almost no shift in pks^- co-cultures. Figure 4.1D shows the percentage of DNA damage positive cells as a function of all tested incubation times. Reporter activity increased within three hours of co-culturing and was maintained for at least two days in 30-50% of cells. In contrast, less than 5% of reporter cells were positive when co-cultured with control pks^- cells. We noted that both pks^+ and pks^- co-cultures showed increased reporter activity at very early time points, likely due to the mechanical pressure experienced when pelleting the cells. The results of our experiments agree with the time frame of the colibactin-triggered DNA damage response reported in other bacteria (Wong et al. 2022) and human cells (Bossuet-Greif et al. 2018; Nougayrède et al. 2006; Reuter et al. 2018; Lucas et al. 2020; Secher et al. 2013).

4.5.2 Colibactin induces DNA damage in distant cells

Once we established that colibactin induces detectable DNA damage within a few hours, we leveraged our fluorescent reporter to evaluate the spatial dynamics of toxicity. In this setup, we followed our DNA damage fluorescent reporter in a lawn of reporter cells surrounding a colony of colibactin producers (Figure 4.3A). We used a fluorescent microscope to quantify the range of colibactin influence around the pks^+ colony. Figure 4.3B shows representative microscopy images from these experiments over three selected time points. As the images show, over

time we detected increased fluorescence in the reporter channel and a seemingly increase in the range of colibactin influence. However, we also detected increases in the CFP and mCherry channels over time, corresponding to an overall increase in cell number on the plates. We therefore had to devise a metric for quantifying the influence range that can account for overall increase in fluorescence.

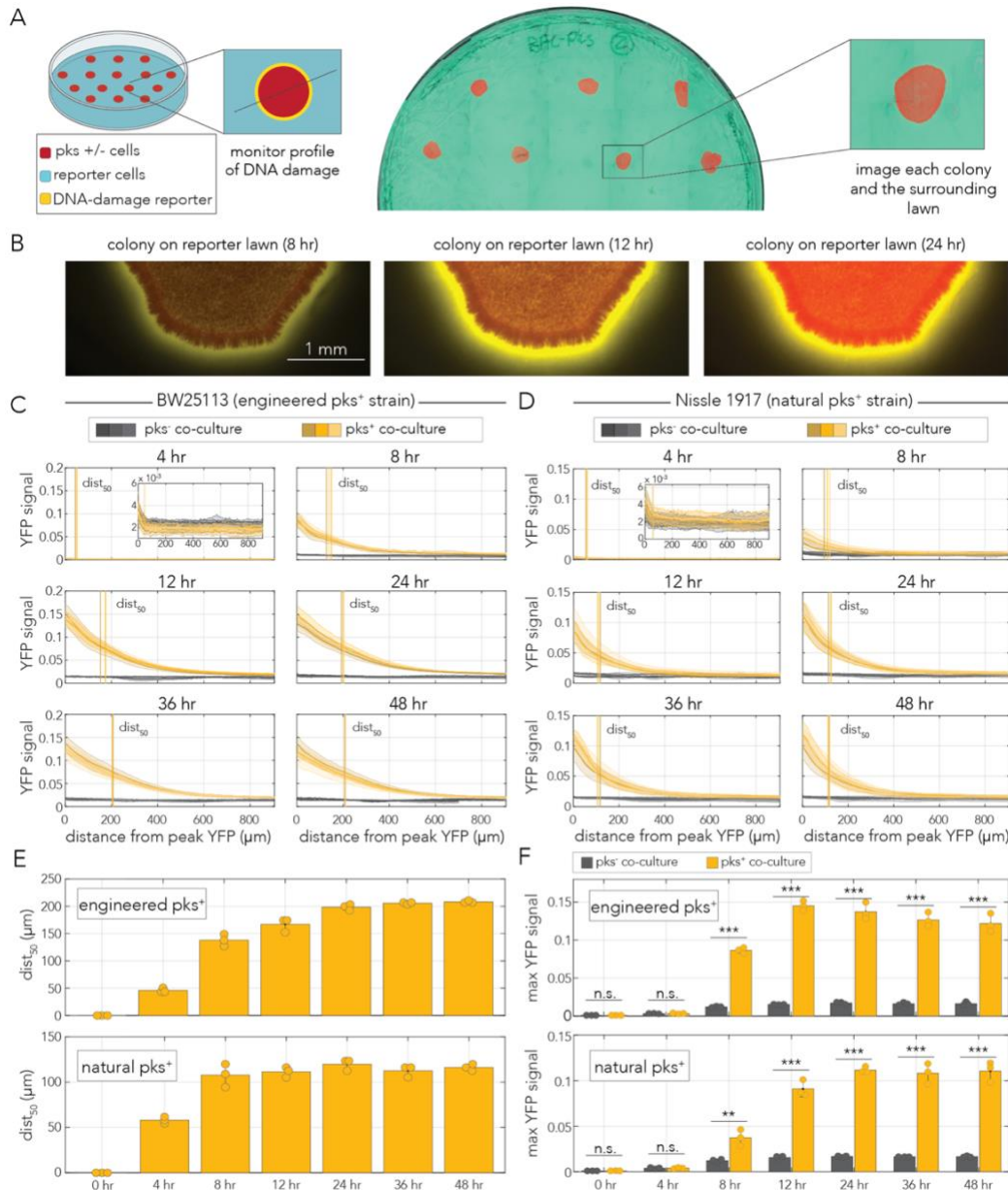


Figure 4.3 Colibactin induces DNA damage in distant cells.

(A) Outline of the experimental setup co-culturing pks^+ or pks^- cells on a lawn of reporter cells. Example image of a reporter lawn expressing CFP with colonies of mCherry-tagged producer cells. (B) Representative microscopy images of a pks^+ colony tagged with mCherry surrounded by a lawn of reporter cells showing activation of the DNA damage response reporter over time. (C-D) YFP signal decay curves over time in reporter cells co-cultured with the engineered pks^+ strain (C) or the natural pks^+ strain (D). The decay curve begins at the peak YFP signal for each technical replicate ($N = 10-15$). Thick lines represent the mean signal intensity for each biological replicate ($n = 3$) and shaded areas represent standard deviation. Vertical lines mark the distance from peak YFP signal at which half of the maximal response was observed for each biological replicate (dist_{50}). Each biological replicate is colored a different shade of yellow (pks^+) or gray (pks^-). (E) Mean distance to half of the maximum YFP signal over time in experiments with engineered pks^+ cells (top) or with experiments with natural pks^+ co-cultures (bottom). Error bars represent standard deviation and points represent the mean of each biological replicate ($n = 3$). (F) Mean maximum YFP signal over time for engineered pks^+ co-cultures (top) and natural pks^+ co-cultures (bottom). Control pks^- co-cultures are shown in gray and the pks^+ are shown in yellow. Error bars represent standard deviation and points represent the mean of each biological replicate ($n = 3$). (** $p < 0.01$, *** $p < 0.001$).

We quantified DNA damage reporter activity along a cross-section that passed from the center of the producer colony (inset in Figure 4.3A). Profiles of YFP intensity in these cross-sections ranging from the edge of the producer colony and into the lawn of reporter cells reflected the DNA damage signal decay curve. Figure 4.3C shows the decay curve averaged across 10-15 colonies. The panels represent curves observed at different time points and the individual lines show results from three independent biological replicates. As the panels show, we observed noticeable decay profiles within eight hours of co-culture and an increase in maximal YFP signal in prolonged incubation periods. In the longest incubation periods, elevated YFP signal was clearly noticeable even 500 μm into the reporter lawn. To extract a metric of spatial penetrance that accounts for overall increase in fluorescence, we calculated the distance to 50% of the

maximal signal (termed dist_{50}). At saturation, the dist_{50} stabilized around 200 μm (vertical lines, Figure 4.3C). To control for potential artifacts in colibactin delivery due to heterologous expression of the pathogenicity island in engineered cells, we repeated the assay with the Nissle 1917 strain that naturally produces the toxin (Figure 4.3D). In these experiments we still clearly observed distant DNA damage, although the range of influence was shorter, with dist_{50} stabilizing around 100 μm .

To summarize and compare the spatiotemporal dynamics in the engineered and naturally producing strains, we examined measurements of two key parameters over time. Figure 4.2E shows dist_{50} in the two strains. In the engineered strain, dist_{50} plateaued after 24 hours and saturated at 208 μm while in the natural producer dist_{50} plateaued already after 8 hours and saturated at 116 μm . The earlier saturation around the natural producer likely arises from lower colibactin expression level in this genetic background. Compatible with this explanation, is the lower signal intensity found at the border of the lawn and the colibactin-producing colony (Figure 4.3F). At saturation, the maximum YFP signal was higher around the engineered producer by almost 50%.

4.5.3 Colibactin-induced DNA damage is independent of cell contact

Given our observation that colibactin induces DNA damage hundreds of microns away from the pks^+ colony (Figure 4.3C-E), we hypothesized that colibactin may induce DNA damage in neighboring, yet clearly separated, colonies. Such an

observation will refute the possibility that DNA damage depends on communication between contacting cells. We therefore designed an additional assay for evaluating DNA damage in both contacting and non-contacting colonies, separated by different distances, on an agar plate. Figure 4.4A shows an overview of the experimental design: a co-culture of producer and reporter cells were spread on agar plates and allowed to form colonies overnight before visualizing them with a fluorescent microscope. To quantitatively characterize the DNA damage response, we measured the reporter activity profile along a cross-section that passes through the centers of a producer and reporter colony pair (inset in Figure 4.4A).

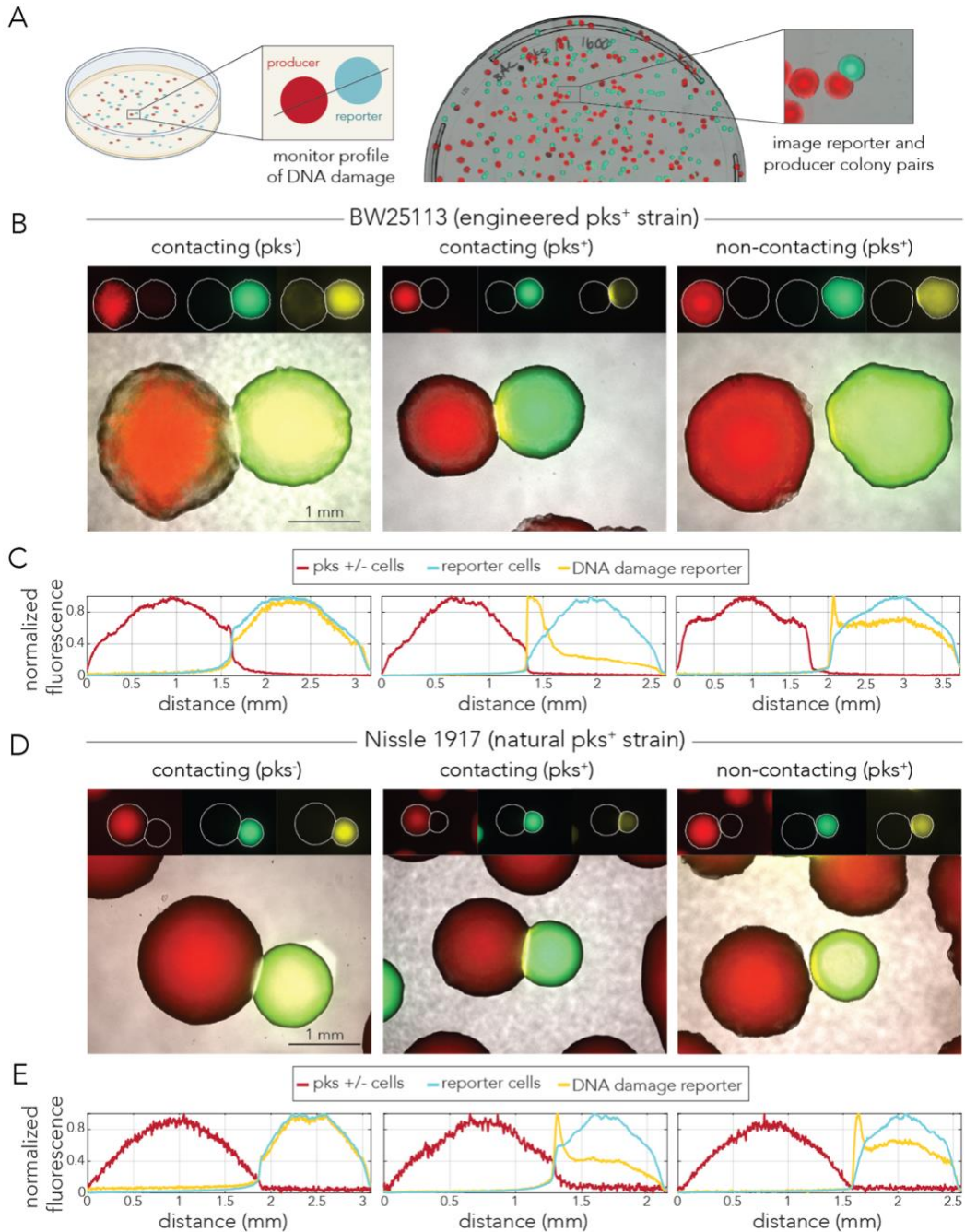


Figure 4.4 Colibactin-induced DNA damage is cell-contact independent.

(A) Outline of the co-culture colony assay. Representative image of a co-culture plate with mixed colonies. (B) Representative microscopy images of reporter colonies (CFP/YFP) next to engineered (BAC) pks^+ or pks^- colonies (mCherry). The signal for each fluorescent tag was min-max scaled in each representative image. In reporter colonies neighboring pks^+ colonies, YFP signal peaks in the region closest to the colibactin-producing colony. (C) Min-max scaled fluorescent cross-section profiles for representative colonies shown in B. (D) Representative microscopy images of reporter colonies (CFP/YFP) next to natural (Nissle 1917)

pks⁺ and pks⁻ colonies (mCherry). The signal for each fluorescent tag was min-max scaled in each representative image. In reporter colonies neighboring pks⁺ colonies, YFP signal peaks in the region closest to the colibactin-producing colony. (E) Min-max scaled fluorescent cross-section profiles for representative colonies shown in D.

Figure 4.4B shows representative microscopy images from the co-culture experiment with an engineered colibactin producer. As expected, we did not observe any skew in the YFP signal when the reporter colony was next to a non-producing colony (Figure 4.4B, left panel). Quantification of fluorescence across all channels, shown in the left panel of Figure 4.4C, supports this observation. The profiles show that the CFP signal, which corresponds to the position and density of the reporter colony, overlaps with the YFP signal. In contrast, reporter colonies contacting pks⁺ cells showed a high intensity YFP signal along a narrow region of contact (Figure 4.4B, middle panel) and quantification of YFP signal showed a clear positional skew (Figure 4.4C, middle panel). Lastly, we also observed a weaker, yet highly reproducible, increase in YFP signal in non-contacting colonies that were in proximity (Figure 4.4B, right panel). This observation was supported by the quantification of the YFP signal (Figure 4.4C, right panel). To control for potential artifacts in colibactin delivery due to heterologous expression of the pathogenicity island, we repeated the assay with the Nissle 1917 strain. Representative microscopy images and their matching fluorescence profiles are presented in Figures 4.4D and 4.4E, respectively. Similarly to the engineered strain, here we also observed a spike in YFP intensity along the edge of both contacting and non-contacting colonies.

We quantified the strength of colibactin-induced DNA damage as a function of the distance from the colibactin-producing colony. Figure 4.5A shows the YFP signal averaged across dozens of colonies from three biological replicates for co-cultures with the engineered colibactin producing strain. As expected, YFP signal remained at a low and flat baseline in the reporter colonies contacting pks^- colonies (Figure 4.5A, left panel). In contrast, YFP signal spiked in reporter colonies contacting pks^+ colonies. To quantify the colibactin influence range, we again used the dist_{50} metric. The average dist_{50} was highly reproducible in biological replicates and averaged at 117 μm from the edge of the producer colony (Figure 4.5C). In non-contacting reporter colonies (separated by at least 10 μm), we also saw clear and reproducible signal decay curves (Figure 4.5A, right panel). For these colonies, we inferred the distance that matches 50% of the maximum signal observed in contacting colonies. The average dist_{50} was highly reproducible in biological replicates and averaged at 113 μm (Figure 4.5C). A two tailed t-test rejected the hypothesis that the average dist_{50} is different for contacting and non-contacting colonies ($p\text{-val} = 0.81$).

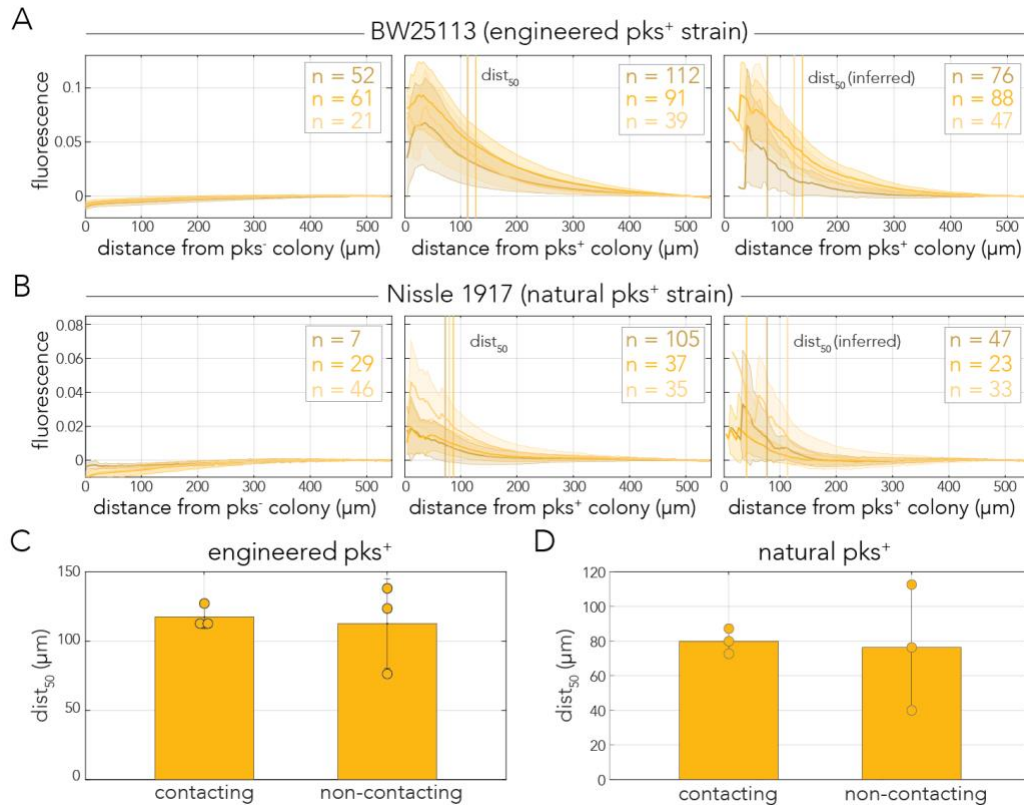


Figure 4.5 Colibactin-induced DNA damage decay similarly with distance in contacting and non-contacting colonies.

(A) YFP decay profiles of reporter colonies plated with engineered cells across three conditions: contacting pks⁻, contacting pks⁺, non-contacting pks⁺. (B) YFP decay profiles of reporter colonies plated with Nissle 1917 cells across the same three conditions. (A, B) YFP signal was aligned to the edge of the mCherry colony and averaged across dozens of colonies in the same plate (n shows the number of colonies monitored). Shaded area marks the standard deviation for each biological replicate. Vertical lines indicate the decay dist₅₀ (50% signal) for each biological replicate. (C) Decay dist₅₀ calculated and averaged averaged across biological replicates for the engineered strain and (D) for the natural strain dist₅₀ in non-contacting colonies is inferred using the averaged YFP intensity measured at dist₅₀ for contacting colonies. Error bars represent standard deviation.

We repeated the colony-based assay to examine signal decay in colonies

neighboring natural colibactin-producing cells. Figure 4.5B shows the

characteristic decay profiles of reporter colonies neighboring Nissle 1917 cells

from three biological replicates. As expected, we observed an overall weaker

reporter signal, compared to the engineered strain. The average dist₅₀ was highly

reproducible in biological replicates of contacting colonies and averaged at 80 μm from the edge of the producer colony (Figure 4.5D). In non-contacting colonies we observed a higher variation across biological replicates, likely due to the lower level of signal, with an inferred dist_{50} of 76 μm . A two tailed t-test rejected the hypothesis that the average dist_{50} is different for contacting and non-contacting colonies ($p\text{-val} = 0.87$).

Lastly, we validated that non-contacting colonies are truly separated from one another and that we are not overlooking pks^+ bacteria that may have swarmed to the visible reporter colony. We reasoned that if swarming cells were sparse, they may go undetected by low magnification microscopy. We used a fine needle to sample the edge of twelve reporter colonies that showed elevated YFP expression and were separated by 60-300 μm from colibactin-producing colonies. We then resuspended these samples and plated them (Figure 4.6). We examined plates that had thousands of colonies, with a median of approximately 7500 colonies, for CFP and mCherry signal. In all experiments we exclusively observed CFP-tagged colonies (Table 4.1). We can therefore conclude swarming pks^+ cells are either absent or extremely rare in the edge of our separated reporter colonies. In summary, our colony-based assay reveals that colibactin-induced damage is evident across clearly non-contacting colonies. Moreover, signal intensity as a function of distance revealed that the signal decay profiles were indistinguishable between contacting and non-contacting colonies.

Table 4.1 Strain purity in non-contacting colonies

pks+ strain	colony pair	distance between colonies (um)	# mCherry colonies	total colonies on plate	CFUs picked
engineered (BAC)	1	211.0878559	0	6713	472000
engineered (BAC)	2	65.39144619	0	7617	948000
engineered (BAC)	3	196.5908064	0	7315	1898400
engineered (BAC)	4	218.0115838	0	6675	1424000
engineered (BAC)	5	141.7858862	0	10387	1284800
engineered (BAC)	6	250.630535	0	9619	1169600
engineered (BAC)	7	316.371115	0	10171	796000
natural (EcN)	1	109.1133998	0	7568	872000
natural (EcN)	2	61.83338299	0	7609	896000
natural (EcN)	3	207.0858625	0	4067	1300800
natural (EcN)	4	130.9372108	0	3662	668800
natural (EcN)	5	119.9792786	0	6802	672000

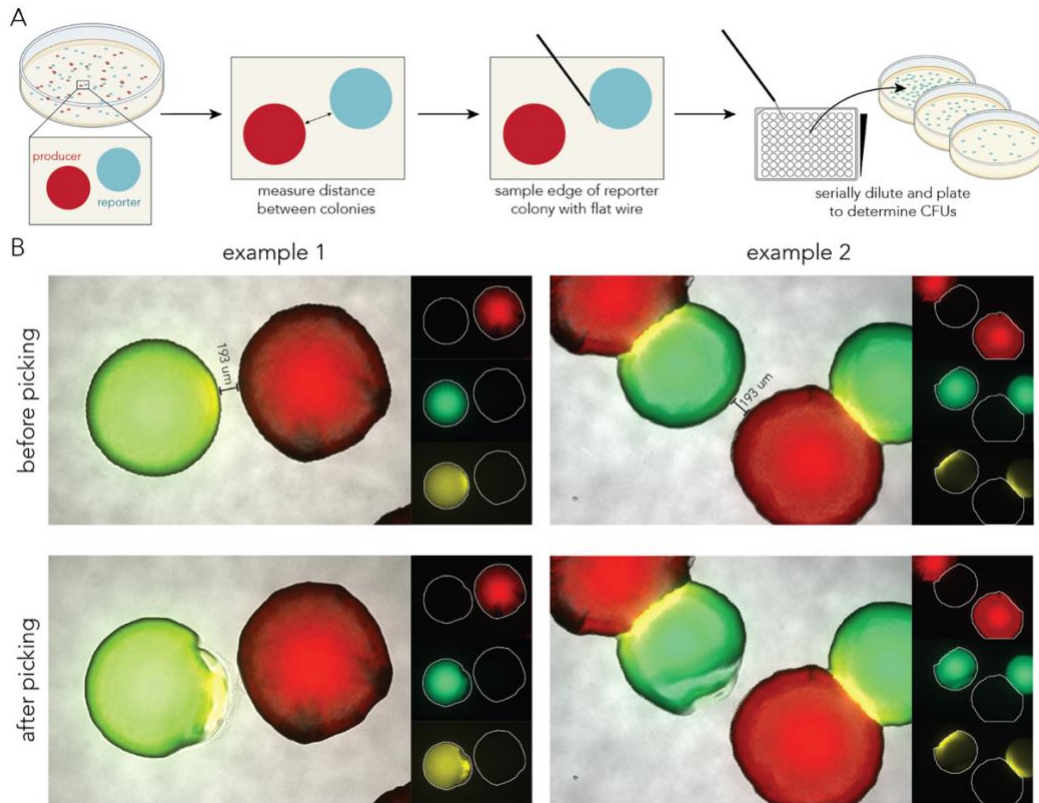


Figure 4.6 Validation of strain purity in the edge of non-contacting colonies

(A) Outline of the sampling and plating approach used to test for strain purity: non-contacting reporter and pk_s^+ colony pairs were identified and the distance between them was measured. Then, the reporter colony was sampled with a fine needle and the sample was suspended in buffer for serial dilution and plating. After overnight growth, plates with highly dense yet separate colonies were inspected with a fluorescent microscope to evaluate strain purity (median colony number was 7,500). (B) Representative images of two non-contacting colony pairs before and after fine needle sampling of the reporter colony. After fine needle sampling of the reporter colony, we confirmed that the pk_s^+ colony remained undisturbed by microscopy imaging (lower panels show only reporter colonies are disrupted while pk_s^+ remained unchanged).

4.6 Discussion

Colibactin-producing bacteria are not uncommon in the gut microbiome of healthy humans and their increased prevalence is evident in multiple human diseases (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al.

2017). Compelling evidence from colorectal tumors strongly supports the premise that colibactin acts as a tumorigenic mutagen (Dziubańska-Kusibab et al. 2020; Pleguezuelos-Manzano et al. 2020). While the clinical relevance of colibactin-induced damage has motivated intense research in mammalian cell cultures, it has left fundamental questions underexplored in bacteria. Addressing these open questions is important for interpreting colibactin's impact on other members of the host microbiome and potentially also for a deeper understanding of colibactin-host interactions. Here we focused on the spatiotemporal dynamics of colibactin-induced DNA damage in bacteria to directly examine the widely accepted premise of cell-contact dependence (Nougayrède et al. 2006; Reuter et al. 2018; Silpe et al. 2022; Bossuet-Greif et al. 2018; Chen et al. 2022).

We monitored the dynamics of colibactin-induced DNA damage by cloning a transcriptional fluorescent reporter that tracks expression of the *recA* gene, a key factor in the homologous recombination DNA repair pathway (Vollmer et al. 1997). By combining this reporter with fluorescent tags that uniquely marked colibactin producers and target cells, we were able to closely track the spatiotemporal dynamics of the DNA damage response in mixed-cell populations. This approach validated previous reports in both bacteria (Wong et al. 2022) and human cells (Bossuet-Greif et al. 2018; Nougayrède et al. 2006; Reuter et al. 2018; Lucas et al. 2020; Secher et al. 2013) by showing that DNA damage is already detectable within several hours of colibactin exposure (Figure 4.1D). However, results from co-culture populations also revealed a seeming discrepancy

with previous works by revealing that DNA damage is detectable hundreds of microns away from colibactin-producing cells (Figure 4.2C-E). The observation of contact independence was reinforced by our colony-based assay that showed that the DNA damage response is triggered in reporter colonies that are clearly separated from colibactin-producing colonies (Figures 4.3C and 4.3E). Lastly, quantification of signal intensity as a function of distance revealed that the signal decay profiles were indistinguishable between contacting and non-contacting colonies (Figures 4.4B and 4.4E). This observation indicated that cell contact does not alter the amount of DNA damage targeted cells experience beyond what is expected by proximity alone.

Taken together, our results establish that colibactin-induced damage in bacteria is cell-contact independent. Since we observed similar effects, albeit weaker, with a strain that naturally produces colibactin, contact independence is not an inadvertent artifact of the high heterologous expression in genetically engineered cells. Importantly, our conclusion contests the premise that cell-cell contact is required for colibactin toxicity in bacteria (Silpe et al. 2022; Wong et al. 2022; Chen et al. 2022). This seeming contradiction can be rationalized by the different experimental methods that we and others used. Previous works inferred contact dependence from experiments that separated producers and target cells with a small pore membrane in a transwell (Silpe et al. 2022) or by measuring viability in mixed populations growing in a colony compared to a planktonic state (Wong et al. 2022). Given our conclusion on the effective distance, it seems likely that

previous experimental setups were not adequately sensitive for detecting DNA damage at very close proximity. Interestingly, since the conclusion favoring the requirement for cell contact in mammalian cells also relied on a transwell-based assay (Nougayrède et al. 2006; Bossuet-Greif et al. 2018), contact-independent interactions taking place at closer ranges may have gone unnoticed. Our work therefore suggests that the premise of contact-dependence should also be reevaluated in mammalian cells with alternative, and highly sensitive, methods.

4.7 Methods

Table 4.2 Bacteria strains used in this study

Strain	Nickname	Source	Use
BW25113 $\Delta gspI::kan$ pRecA**	Reporter strain	This study	DNA damage flow cytometry and microscopy
BW25113 pBeloBAC11+pks- mCherry*	Engineered pks ⁺ mCherry	This study	DNA damage flow cytometry and microscopy
BW25113 pBeloBAC11- mCherry*	Engineered pks ⁻ mCherry	This study	DNA damage flow cytometry and microscopy
Nissle 1917 (EcN) $\Delta mchDEF::kan$ mCherry	Nissle 1917 pks ⁺ strain (microcin deletion)	This study	DNA damage microscopy
Nissle 1917 (EcN) $\Delta mchDEF::kan$ $\Delta clbN::chl$ mCherry	Nissle 1917 pks ⁻ strain (microcin deletion)	This study	DNA damage microscopy

*parental BW25113 pBeloBAC11 and BW25113 pBeloBAC11+pks strains were received from the Balskus lab (Silpe et al. 2022).

**parental BW25113 $\Delta gspI::kan$ strain was from the Keio collection (Baba et al. 2006).

4.7.1 Media and growth conditions

All experiments were performed in either Lysogeny Broth (LB) or minimal synthetic media (M9 salts supplemented with 0.4% glucose, 2 mM MgSO₄, 0.1

mM CaCl₂, 0.2% ampicillin). Overnight cultures for all experiments were grown at 37°C with 200 rpm orbital shaking. During the overnight growth of antibiotic-resistant strains, we added antibiotics at the following concentrations: 50 µg/mL spectinomycin, 50 µg/mL kanamycin, 25 µg/mL chloramphenicol, and 50 µg/mL carbenicillin. Agar-based experiments were conducted on M9 1.5% agar.

4.7.2 Cloning deletion strains

The two microcins MccM and H47 were deleted from Nissle 1917 using lambda red recombination. The insert had 40 bp homology arms to the upstream and downstream genomic regions of the *mchDEF* genes and a kanamycin resistance cassette. *clbN* was deleted from Nissle 1917 using lambda red recombination with an insert containing 40 bp homology arms to the upstream and downstream genomic regions and a chloramphenicol resistance cassette.

4.7.3 Fluorescent reporter plasmids

We cloned the DNA damage reporter plasmid with the Gibson assembly method (Gibson et al. 2009) using In-Fusion Snap Assembly Master Mix (Takara, cat# 638947). In a single assembly reaction, we integrated a YFP and the *recA* promoter into a plasmid backbone containing a spectinomycin resistance cassette and CFP. When amplifying the backbone, we also replaced the CFP promoter with a 48 bp EM7 promoter that was encoded on the amplification primer. The 81 bp *recA* promoter was amplified from the BW25113 genome (the promoter region was defined according to previous work (Salgado et al. 2023; Pagès et al. 2003;

Stohl et al. 2003)). The final plasmid was a low-copy plasmid with the sc101 origin and spectinomycin resistance. Plasmid assembly was validated with Sanger sequencing spanning the integration sites.

The BAC and Nissle 1917 pks⁺ and pks⁻ strains were cloned to express an mCherry plasmid. The plasmid was high-copy with a colE1 origin and an ampicillin resistance cassette. The mCherry is expressed under the constitutive UV8 promoter.

4.7.4 Monitoring DNA damage reporter with flow cytometry

Cultures of the reporter strain and engineered pks⁺ and pks⁻ strains containing mCherry expressing plasmids were grown overnight in LB. 1 mL of overnight culture was washed three times in PBS and resuspended in M9 media. Cultures were then diluted 1:200 and allowed to grow for 12 hours. Following growth, optical density at 600 nm (OD₆₀₀) was measured and cultures were diluted to OD₆₀₀=0.1. The reporter culture was then diluted 1:10 into M9 media that contained either the engineered pks⁺ or pks⁻ strain at the same density. The co-culture was maintained in a 96 deep-well plate (Eppendorf, cat# 2231000920) in a final volume of 500 µL. The co-cultures were pelleted at 4000 g for 6 minutes and were then incubated at 37°C without shaking for up to 48 hours. At different pre-determined time intervals, specific wells were resuspended and transferred to a new 96 deep-well plate. Co-cultures were then pelleted at 4000 g for 6 minutes and supernatant aspirated. Cells were fixed with 3.7% formaldehyde for 15

minutes at room temperature. Cells were washed with PBS twice and resuspended in 300 μ L PBS. Fixed cells were stored at 4°C for up to three days, then analyzed by flow cytometry. Flow cytometry was performed with BD LSRFortessa with a high throughput sampler. CFP was measured with a 405 nm laser with a 525/50 nm filter and 505 nm long pass filter. YFP was measured with a 488 nm laser with a 530/30 nm filter and 505 nm long pass filter. mCherry fluorescence was measured with a 561 nm filter with a 610/20 nm filter and 600 nm long pass filter.

We analyzed flow cytometry data with a custom MATLAB (MathWorks) script. For gating purposes, unstained, single color, and sample treatment cell populations were overlaid in plots. To separate cell populations from debris, we first gated events using forward scatter and side scatter areas. The filtered events were then further gated to identify single cells with forward scatter area and forward scatter height. Single cells were then gated on each fluorescent channel (by fluorescent channel area). We used a conservative cutoff to classify reporter cells that were positive for the DNA damage response (by YFP signal). We relied on the YFP intensity distribution of the reporter cells at the start of the experiment (t=0 hours) and set the cutoff to be the mean intensity plus three standard deviations.

4.7.5 Monitoring DNA damage reporter by microscopy in a lawn co-culture

Cultures of the reporter strain, the engineered pks^+ and pks^- strains, and Nissle 1917 pks^+ and pks^- strains (with a microcin deletion, as microcins are also toxic to

neighboring cells) were grown overnight in LB with antibiotics. The density of overnight cultures was measured using OD₆₀₀. Strains were then washed twice in PBS and then the reporter strain was diluted to an OD₆₀₀ of 4 and the pks⁺ and pks⁻ strains were concentrated to an OD₆₀₀ of 10. 150 μL of reporter strain was spread on M9 agar plates and allowed to dry. The pks⁺ and pks⁻ strains were spotted in a 1 μL volume and immediately imaged for a 0-hour time point. Plates were incubated at 37°C until each subsequent time point up to 48 hours. Colonies and the surrounding lawn were imaged at 2.5x magnification using CFP (475nm), YFP (524nm), mCherry (610nm), and brightfield channels on a Zeiss Axio Observer.Z1 epifluorescence microscope.

Microscopy image analysis was performed with custom scripts in MATLAB (MathWorks). Briefly, a user marked a line from the center of the producer colony extending into the reporter lawn using the images obtained from the bright-field and CFP channels. The pixel intensity for each fluorescent channel was determined along this cross-section line. Signal decay away from a colibactin producing colony was measured from the peak YFP signal into the reporter lawn. Maximum YFP signal intensity was reported from the peak of the YFP signal. We calculated dist₅₀, the distance at which we observed 50% of the maximal response, for contacting colonies by first smoothing each YFP signal decay with a moving window of 20. Decays were smoothed to control for noise in the baseline of the YFP signal, especially at early time points where the overall signal was low. We then found the fluorescent value halfway between the max YFP signal and the

baseline signal (YFP₅₀). The baseline signal was determined by calculating the average YFP signal of the furthest 20 pixels from the pks⁺ colony. We then found the distance along the decay profile that corresponds to YFP₅₀.

4.7.6 Monitoring DNA damage reporter by microscopy in single colonies

Cultures of the reporter strain, the engineered pks⁺ and pks⁻ strains, and Nissle 1917 pks⁺ and pks⁻ strains (with a microcin deletion, as microcins are also toxic to neighboring cells) were grown overnight in LB with antibiotics. The density of overnight cultures was measured using OD₆₀₀ and were mixed at a ratio of 5:1 or 1:1 (producers to reporters). Mixed co-cultures were then diluted to multiple concentrations and spread on M9 agar plates with glass beads to achieve a range of 400-1,600 colonies per plate. After 24 hours of incubation at 37°C, plates were imaged with a Zeiss Axio Observer.Z1 epifluorescence microscope. Colonies were imaged at 2.5x magnification using CFP (475nm), YFP (524nm), mCherry (610nm), and brightfield channels.

Microscopy image analysis was performed with custom scripts in MATLAB (MathWorks). Briefly, a user marked a line crossing two neighboring colonies and additional lines marking the colony edges on the image obtained from the bright-field and CFP channels. The pixel intensity for each fluorescent channel was determined along this cross-section line. Signal decay away from a colibactin producing colony was measured from the peak YFP signal (at the reporter colony edge) and into the reporter colony itself. For non-contacting colonies, values at

positions between the two colonies were ignored. Baseline colony YFP expression was determined by calculating the mean intensity of the furthest 20 pixels from the pks colony in the decay curve. This value was subtracted from each decay curve to control for differences in baseline YFP expression in each colony. The distance between each colony pair was calculated by the user marked colony edges.

We calculated dist_{50} for contacting colonies by first finding the fluorescent value halfway between the max YFP signal and the baseline signal (YFP_{50}). We then found the distance along the decay profile that corresponds to YFP_{50} . The dist_{50} for non-contacting colonies was inferred by the YFP_{50} value measured earlier for contacting colonies (averaged across all replicates). This was to account for the fact that non-contacting colony decay curves were not peaking at the maximum YFP reporter response.

4.6.7 Colony cross contamination test

We tested strain purity in the edge of non-contacting colonies to validate that increased YFP intensity is not attributed to swarming pks^+ cells that were undetectable with the low magnification microscope. This experiment setup was identical to the single colony microscopy assay described above. Non-contacting co-culture colonies were identified on agar plates and the distance separating them was calculated by image analysis as described previously. The colonies were observed under a LEICA S6 E microscope at 2x magnification and the edge of the

reporter colony was picked with a flattened platinum wire (forming a fine needle). The collected cells were inoculated into PBS from the wire picker for plating. After picking, we confirmed that only the reporter colony was disturbed at 2.5x magnification on a fluorescent microscope. We excluded any colony pairs where the pks⁺ colony appeared disrupted upon visual inspection at 2.5x on a fluorescent microscope. The suspended reporter colony samples were serially diluted 1:1 for four dilutions (in 1 mL PBS total) before 100 μ L from each dilution was plated on LB agar plates. The plates were cultured overnight at 37°C and imaged the following morning with a Canon EOS Rebel T3i camera. Colony counts were calculated with FIJI and CFUs were back-calculated based on the dilution factor. Plates were imaged on a microscope with CFP (475nm) and mCherry (610nm) channels to confirm the presence of only CFP-tagged colonies.

4.7.7 Code availability

MATLAB codes used in this article are deposited to GitHub (DOI: 10.5281/zenodo.11659009).

Chapter 5: Discussion

5.1 Summary

Colibactin has been predominantly studied in the context of colon cancer and the prevalence of colibactin-producing bacteria in certain disease backgrounds, including inflammatory bowel disease (Buc et al. 2013; Dejea et al. 2018; Iyadorai et al. 2020; Eklöf et al. 2017). While much progress has been made in understanding the role colibactin plays in tumorigenesis, the DNA repair pathways in host cells responding to colibactin-induced damage have not yet been characterized. Nor has the mechanism of cellular entry been identified. Despite its role in tumorigenesis, colibactin likely evolved as a form of interference in bacterial competition, yet its impact on other bacteria has not been thoroughly studied. This thesis work aimed to employ genetic screens in both human host cells and bacteria to systematically determine cellular responses to colibactin toxicity. These screens revealed key DNA repair pathways activated by colibactin-induced damage, including homologous recombination in *E. coli* and Fanconi anemia/ICL repair and fork quality control in humans. However, we were unable to identify a mechanism of cell entry, which has remained elusive since the toxin was first discovered. Our mutation accumulation experiment showed that colibactin binds the same DNA motif in humans and *E. coli*, but that the resulting mutation differs. This difference in mutations could be due to different biases in prokaryotic and eukaryotic cells in erroneous DNA repair. We also identified a genomic footprint in colibactin-producing cells showing that they experience self-

damage. Finally, we found that cell-cell contact is not required for colibactin toxicity in *E. coli*. While this work reduced the gap between our understanding of colibactin toxicity in human cells and *E. coli* cells, there are still unknowns that could be different between the two systems, such as cell entry. This is especially interesting to learn due to the differences in cell membranes between eukaryotes and prokaryotes, with the majority of bacteria also possessing a cell wall, and the presence of a nuclear envelope as an additional barrier between colibactin and DNA in eukaryotes.

5.2 Limitations of this work

There are several limitations to this work that should be taken into consideration and inform future investigations. First, the whole-genome CRISPR screen conducted in Chapter 2 was in a cancer cell background. Thus, mutations exist in certain genes making them over or under active compared to healthy colon epithelial cells. These mutations can impact the extent to which single gene knockouts mediate a phenotype. One concern with the cell line we used, HCT116, is that it is MMR-deficient, so we cannot fully evaluate the role of MMR in responding to colibactin-induced damage. Unfortunately, healthy colon epithelial cell lines are not readily available. An alternative would be to conduct a screen in the RPE1 retinal epithelial cell line that is immortalized with a telomerase reverse transcriptase (Lee et al. 2004), which allows these otherwise healthy cells to be cultured like cancer cell lines.

The genetic screen and mutation accumulation experiments in Chapter 3 establish mechanisms of DNA repair and a mutational signature in *E. coli*. Due to our exclusive use of *E. coli* in our experiments, our conclusions are limited to *E. coli*, but may extend to gram-negative bacteria. While toxicity has been demonstrated in gram-positive bacteria (Chen et al. 2022; Silpe et al. 2022; Wong et al. 2022), mutation analysis has not been conducted. A next step from this thesis work would be to conduct mutation accumulation studies first in a gram-positive species. Should the mutational signature be reproduced in a gram-positive species, we would next recapitulate this result in a bacteria community representative of the human gut. Not only would this establish whether colibactin mutates bacteria DNA in general, it would also provide more data on how colibactin may shape the host gut microbiome by damaging certain species more than others. Additionally, our bioinformatics analysis of self-inflicted DNA damage was limited to *E. coli*, largely due to the predominance of colibactin in the B2 phylogroup of *E. coli*. However, other species, including *Klebsiella pneumoniae* produce the toxin. Dozens of *K. pneumoniae* isolates have been annotated as pks⁺, providing a sample to test for a trinucleotide skew associated with colibactin. This sample is not nearly as large as the *E. coli* dataset we analyzed, but may still shed light on the general trend of trinucleotide skew even if sufficient power is lacking to make a statistical claim.

A final limitation relates to our finding in Chapter 4 that colibactin does not require cell-cell contact in *E. coli*. Our conclusion relies on macroscopic assays

that leave some room for doubt as to whether there is truly no cell-cell contact occurring. First, our assay involving reporter and pks⁺ colonies spread on an agar plate does not involve a physical barrier. We used 1.5% agar, which is dense enough to inhibit swimming and swarming, but does not ensure that cells are fixed in place. We addressed this concern by picking the edge of reporter colonies and plating to determine the frequency of pks⁺ cells. While we did not observe any pks⁺ cells from any of our samples, there remains the possibility that pks⁺ cells existed in our sampling at a rate of less than one pks⁺ cell for every 10,000 reporter cells. Should this be the case, the intensity of reporter activity for the DNA damage response would likely be much lower in non-contacting colonies than what we observed (Figure 4.5). The lack of pks⁺ cells in our colony sampling and the same level of reporter intensity in contacting and non-contacting colonies led us to conclude that colibactin does not require cell-cell contact in *E. coli*, despite the limitations presented. Due to colibactin's instability and inability to be isolated, methods with a barrier between cells are unlikely to be successful, as the cells need to be within several hundred microns of each other. A simpler approach could be accomplished in co-cultures with mammalian cells, where the contact requirement needs to be reevaluated based on our findings. This approach would involve an adapted transwell assay in mammalian cells. Tumor cells could be grown on a cover slip stacked on a platform or additional coverslips to within 200 μm of the transwell insert, allowing bacteria to be within our observed distance but separated by a membrane. This assay would prove challenging in bacteria as they are not adherent, thus a custom device would need to be produced.

5.3 Implications of this work

5.3.1 Potential for host-specific colibactin therapeutics

A major concern with colibactin as a tumorigenic agent is how to prevent the accumulation of mutations in individuals with pks⁺ bacteria in their gut microbiome. Approximately 20-30% of healthy individuals have pks⁺ bacteria in their microbiome (Nougayrède et al. 2006; Dubois et al. 2010; Watanabe et al. 2020; Dejea et al. 2018; Eklöf et al. 2017; Dubinsky et al. 2020), potentially predisposing them to colorectal cancer. One group developed mimics of the biosynthetic colibactin precursor that bind to ClbP, inhibiting the ClbP-mediated final cleavage of colibactin (Volpe et al. 2023). One of these inhibitors showed strong specificity to ClbP when tested in bacteria with serine hydrolases and penicillin binding proteins, which were predicted to be the strongest candidates for off-target effects. While this inhibitor is highly specific to colibactin production and is able to protect host cells from DNA damage, its implications on the gut microbiota have not been evaluated. There is limited data around the role that colibactin may play in modulating the gut microbiota. However, one study of the mouse gut demonstrated that the presence of pks⁺ bacteria altered phyla abundance and overall diversity, specifically that colibactin increased diversity in the mouse gut (Tronnet et al. 2020). This suggests that targeting colibactin biosynthesis with inhibitors, while protecting the host from DNA damage, will have broad and potentially negative implications in the host gut microbiome. To protect the host from DNA damage while still maintaining colibactin toxicity to

other bacteria, we need to identify host-specific mechanisms of colibactin toxicity to develop proper therapeutics.

One of the key questions this thesis work addressed is if colibactin has the same molecular and cellular impacts in bacteria as it does in mammalian cells. We found that colibactin binds A/T rich motifs in *E. coli*, as it does in human cells (Pleguezuelos-Manzano et al. 2020; Dziubańska-Kusibab et al. 2020), but that the resulting mutation differed. This is likely due to biases in the repair processes responding to colibactin-induced damage in each system. We also aimed to identify mechanisms of cellular entry in both human and *E. coli* systems. A difference in cell entry mechanism may provide a host-specific target for colibactin preventative therapeutics. Surprisingly, neither of our genetic screens identified a mechanism of cellular entry. Past works identified colibactin in outer membrane vesicles (Cañas et al. 2016; Fábrega et al. 2017), which alone were able to induce DNA damage in host cells. One study in bacteria found that *N*-myristoyl-D-Asn (NMDA), an abundant pks island metabolite, permeabilizes bacteria membranes, suggesting a potential mechanism of cellular entry (Wong et al. 2022). However, should this be a mechanism of entry, it would likely be unique to bacteria as human cells have NMDA receptors on the cell surface (Kim et al. 2008). We did not detect involvement of any NMDA receptors in our whole-genome CRISPR screen in Chapter 2, which suggests that colibactin does not have a potential interaction with these channels, such as hitchhiking through the channels.

To properly target colibactin toxicity in host cells, further work needs to identify a unique target. Cellular entry is a potential candidate, as we do not yet know how colibactin enters either human or bacteria cells. The difference in cell membrane composition may mean different modes of entry into eukaryotic and prokaryotic cells. NMDA membrane permeabilization as suggested in bacteria is unlikely to be the mechanism in mammalian cells due to the presence of NMDA receptors on the cell surface and the additional barrier of the nuclear envelope. Colibactin is a small molecule, and could possibly passively diffuse across all membranes, independently of NMDA. Should this be the case, an alternative target would need to be identified. Given the similarity between human and *E. coli* cell responses to colibactin toxicity revealed in this thesis, there is the possibility that a host-specific target does not exist. Should this be the case, more work to understand colibactin's role in microbial interactions in the gut microbiome is needed to understand the implications of a broad colibactin inhibitor. A broad colibactin inhibitor would only be successful if gut microbial diversity and abundance can persist at sufficient levels to maintain host health, which would require extensive gut community experiments.

5.3.2 Colibactin-induced DNA damage as a biomarker

Characterizing the gut microbiota in healthy versus disease backgrounds has been a growing field and has produced hypotheses that the gut microbiome can be used as a diagnostic marker for colon cancer. This hypothesis could extend to using colibactin as a marker of colon cancer risk. Several approaches have investigated

the gut microbiome composition and whether pathogenic strains, such as *B. fragilis* and *F. nucleatum*, could be used as diagnostic markers (Nagavardhini and Chittibabu 2022). While many studies have focused on correlated species presence or abundance with disease states, some have turned to the presence of certain genes. One study found that changes in species and genes in the gut microbiome are not necessarily associated with functional changes and that functional features had more accuracy in predicting if samples were controls or from colon cancer patients (Norouzi-Beirami et al. 2020).

The presence of colibactin-producing bacteria in the gut has been proposed as a potential diagnostic marker (Veziant et al. 2021; Strakova et al. 2021; Gaab et al. 2023). Our identification of a mutational signature in colibactin-exposed *E. coli* and a genomic footprint in colibactin producing bacteria suggests a new approach to using colibactin as a diagnostic marker. Simple screening for pks⁺ bacteria would include a PCR for at least one pks island gene. From there, a conserved region in bacterial genomes that contains the A/T rich motif could be periodically sequenced to look for the frequency of mutation, measured as the ratio of ATA/TAT over AAA/TTT. As mutations accumulate, the patient would be recommended for more invasive screening for colon cancer, as increased mutations in the gut bacteria would be indicative of increased mutations in the colon epithelial cells.

5.3.3 Bacterial metabolites in cancer

The role of bacteria and their metabolites in disease is a thoroughly investigated field, but also a vast field with many unknowns remaining. In addition to the need for many population studies to make accurate predictions about the role microbes play in disease, many bacteria-secreted metabolites have unknown functions.

Colibactin was discovered less than 20 years ago, and only for its growth arresting properties in co-cultures with eukaryotes (Nougayrède et al. 2006), and created a niche field dedicated to understanding its role in cancer and the gut microbiome.

This field still has open questions, some of which this thesis work addressed, but the toxin serves as an example that 1) there is much to learn still about the gut microbiome and 2) gaining a thorough understanding of even a single metabolite can take decades.

There are many types of bacteria products that may interact with the host, like colibactin does. Many recent studies have focused on the roles of short-chain fatty acids and secondary bile acids in the relationship between diet, the gut microbiome, and host diseases (Louis et al. 2014). The most prominent phyla in the human gut are *Firmicutes* and *Bacteroidota* (Eckburg et al. 2005; Turnbaugh et al. 2009; Arumugam et al. 2011). Many *Firmicutes* species are specialized in carbohydrate metabolism and their abundance changes quickly in response to dietary changes to carbohydrate intake (Louis et al. 2014; H et al. 2007; Walker et al. 2011). For example, certain butyrate-producing *Firmicutes* species decrease in abundance in response to a low-carb diet. However, butyrate, along with

propionate, is implicated in suppressing inflammation and cancer in the gut (Louis et al. 2014; McNabney and Henagan 2017). Gut bacteria can also metabolize phytochemicals into compounds that can inhibit pro-inflammatory factors (Cardona et al. 2013). Conversely, bacteria metabolites can also promote inflammation and cancer. Polyamines can be toxic in high levels and some bacteria, such as *B. fragilis*, upregulate polyamine production in host cells (Goodwin et al. 2011). Sulfide is genotoxic and is produced by bacteria, such as *Desulfovibrio* species, that metabolize sulfate taken in through the host diet (Marquet et al. 2009; Louis et al. 2014). Bacteria are also able to metabolize host-produced bile acids to secondary bile acids. One secondary bile acid, deoxycholic acid, can cause genomic instability by oxidative damage in colon epithelial cells (Jenkins et al. 2008; Glinghammar et al. 2002; Huo et al. 2011), while lithocholic acid can damage the epithelial lining via reactive oxygen species production, which leads to apoptosis resistance and increased proliferation (Vogel et al. 2012; Schwarcz et al. 2024).

The role that bacteria metabolites can play in host health and disease is extensive. The examples provided hardly scratch the surface of what we already know, but also provide insight into how much work needs to be done for us to fully understand negative interactions between the host and its microbiome. Colibactin is one example of how simple co-cultures of patient isolates with cell lines demonstrated a growth arrest phenotype, leading to the finding that a novel toxin, colibactin, was inducing this phenotype. This approach of screening isolates for

growth inhibition or DNA damage in human cell lines may identify carcinogenic compounds in the gut microbiome and complement the population studies that are ongoing to identify species and genes associated with cancer. A limitation to population studies is the need for large sample sizes and sequencing depth. Thus, screening bacteria isolates from only a few individuals for growth arrest or DNA damage in human cells could provide a faster approach to identifying novel carcinogens. In fact, some research groups are using high-throughput methods to associate microbial metabolites with other host diseases, such as cardiovascular disease (Li et al. 2024). The experimental approaches used in this thesis provide an approach for characterizing novel microbial metabolites with respect to their impacts on the host and if these metabolites also interact with other gut microbes.

5.4 Conclusions

The work in this thesis highlights the utility of genetic screens in model organisms, such as *E. coli*, and CRISPR screens in human cell lines to systematically evaluate underlying mechanisms of novel bacteria products. These approaches allow for a complementary understanding of the impact these products have on host disease and how they impact the gut microbiome, especially in the case of bacterial competition. Host-microbiome interactions are complex, but these methods can provide a baseline for functionally characterizing bacteria products and identifying if the products have host-specific interactions. The data gained from these experiments can further inform therapeutic targets to maximize host protection while minimizing alterations to the gut microbiome. In the case of

colibactin, further work is needed to characterize its impact on the gut microbiome to determine how a broad colibactin inhibitor would alter the gut microbiome and influence host health. Targeted studies also need to be conducted to rule out a mechanism of cellular entry unique to humans. These targeted studies could identify a host-specific therapeutic target that would maintain colibactin's interactions with the gut microbiota.

Chapter 6: Appendix I: Rapid signaling reactivation after targeted BRAF inhibition predicts the proliferation of individual melanoma cells from an isogenic population

6.1 Preface

The work in this appendix is adapted from the following publication:

Khoshkenar, P., Lowry, E. & Mitchell, A. Rapid signaling reactivation after targeted BRAF inhibition predicts the proliferation of individual melanoma cells from an isogenic population. *Sci Rep* **11**, 15473 (2021).

<https://doi.org/10.1038/s41598-021-94941-8>

The author contributions are as follows:

P.K.—Experiments, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology; E.L.—Experiments, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing and Editing; A.M.—Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing and Editing.

I conducted the bulk measurement experiments presented in Figures 6.1 and 6.2.

Payam Khoshkenar conducted single cell measurement experiments presented in Figures 6.3, 6.5, 6.6, 6.7, 6.8.

6.2 Abstract

Cancer cells within tumors display a high degree of phenotypic variability. This variability is thought to allow some of the cells to survive and persist after seemingly effective drug treatments. Studies on vemurafenib, a signaling inhibitor that targets an oncogenic BRAF mutation common in melanoma, suggested that cell-to-cell variation in drug resistance, measured by long-term proliferation, originates from epigenetic differences in gene expression that pre-exist treatment. However, it is still unknown whether reactivation of signaling downstream to the inhibited BRAF, thought to be a key step for resistance, is heterogeneous across cells. While previous studies established that signaling reactivation takes place many hours to days after treatment, they monitored reactivation with bulk-population assays unsuitable for detecting cell-to-cell heterogeneity. We hypothesized that signaling reactivation is heterogeneous and is almost instantaneous for a small subpopulation of resistant cells. We tested this hypothesis by monitoring signaling dynamics at a single-cell resolution and observed that despite highly uniform initial inhibition, roughly 15% of cells reactivated signaling within an hour of treatment. Moreover, by tracking cell lineages over multiple days, we established that these cells indeed proliferated more than neighboring cells, thus establishing that rapid signaling reactivation predicts long-term vemurafenib resistance.

6.3 Introduction

Numerous studies over the past decade revealed extensive diversity between cancer cells within tumors. These cell-to-cell differences, often referred to as heterogeneity, can involve multiple cellular pathways and underly significant phenotypic variation. In some cases, cell-to-cell heterogeneity can engage key oncogenic pathways underlying disease progression or influencing treatment resistance and can therefore pose significant challenges for personalized cancer treatment (Burrell et al. 2013). Cell-to-cell variation is typically thought to originate from genomic instability leading to genetic differences between individual cancer cells within a tumor. However, multiple recent studies have also revealed that non-genetic mechanisms can play a significant role in promoting heterogeneity with potential important implications for correct diagnostics, disease progression and treatment options (Black and McGranahan 2021; Flavahan et al. 2017; Puig et al. 2018; Shaffer et al. 2017; Sharma et al. 2010). In these cases, isogenic cancer cells exist in alternative epigenetic states that arise from transcriptional and post-transcriptional differences between them (Marjanovic et al. 2020; LaFave et al. 2020). Since the underlying mechanisms driving this type of heterogeneity can be independent of genetic differences (mutations) between cells, an individual cell can readily transition between two or more phenotypic states (Shaffer et al. 2017; Sharma et al. 2010).

The research of epigenetic-driven heterogeneity has greatly benefitted from in-vitro experiments in cell-line cultures. Such experiments, especially with clonal cell-lines that were derived from a single ancestor cell, allow controlling for

multiple parameters that are highly variable in the extracellular tumor microenvironment and can also contribute to heterogeneity. Such parameters include uneven nutrient availability, oxygen gradients and spatial pH differences across the tumor environment (Hoefflin et al. 2016; Lloyd et al. 2016; Korenchan and Flavell 2019). Hence, in-vitro experiments allow for elucidation and focus on the inherent phenotypic diversity that is still retained even in a highly regulated and uniform in-vitro environment (Puig et al. 2018; Shaffer et al. 2017; Sharma et al. 2010). Melanoma cell-lines harboring the oncogenic BRAF^{V600E} mutation have become a primary model system for studying adaptive drug resistance and epigenetic-driven heterogeneity (Shaffer et al. 2017; Fallahi-Sichani et al. 2017; Lito et al. 2012; Fallahi-Sichani et al. 2015). Specifically, the BRAF^{V600E} mutation, which is the most frequent BRAF mutation in melanoma, stimulates the constitutive activation of the downstream extracellular signal-regulated kinase (ERK). Constitutive activation of this pro-proliferative kinase in tumor cells, even in the absence of any extracellular stimuli, allows cells to become self-sufficient in growth signals (Cantwell-Dorris et al. 2011). Vemurafenib, the first drug approved for BRAF-mutant cancer, is a selective inhibitor of oncogenic BRAF that offers a significant clinical benefit for patients with metastatic melanoma (Bollag et al. 2012). Vemurafenib acts by selectively binding to the mutated BRAF protein which in turn inhibits the downstream mitogen-activated protein kinase network, including ERK.

While vemurafenib emerges as a highly effective anti-cancer treatment reaching 80% tumor response rate among patients, its potency is typically only transient,

and after initial response, drug resistance arises in most patients (Flaherty et al. 2024). A recent pioneering study that investigated vemurafenib resistance in-vitro in isogenic populations of melanoma cells revealed that drug resistance can arise from a reversible epigenetic state that characterizes a rare subpopulation of cells (Shaffer et al. 2017) (Fig. 6.1A). The study leveraged on cell cultures that were plated sparsely for monitoring the long-term fate of individual cells and revealed that resistance is extremely heterogeneous and is correlated with high expression of multiple resistance marker genes, many of which belong to the gene network responding to extra-cellular growth factors (e.g., the growth factor receptors NGFR and EGFR).

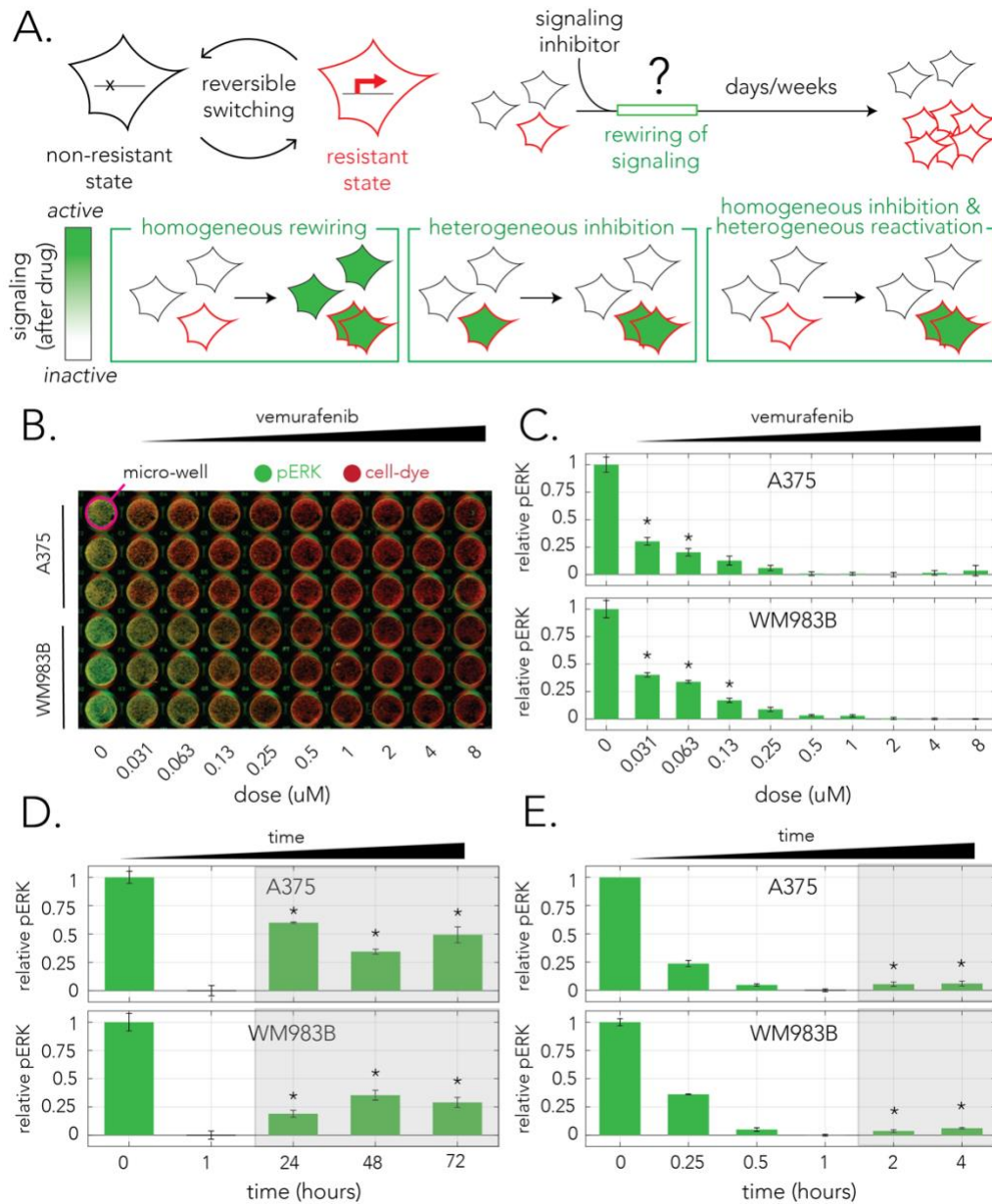


Figure 6.1 Reactivation of ERK signaling after vemurafenib treatment in bulk populations.

(A) The epigenetic-based model for drug resistance and alternative models for signaling rewiring. High expression of a set of genes is thought to underlie a reversible drug-resistant state (upper-left panel). A small subpopulation of drug-resistant cells proliferates more than kin-cells after drug administration, however, the dynamics of signaling rewiring immediately after drug treatment in individual cells remains unknown (upper-right panel). The lower panels show three alternative models for signaling rewiring consistent with measurements from bulk populations. (B) A representative image of the immunofluorescence signal in the in-cell-western (ICW) method used for monitoring signaling activity 1 h after

treatment. Cells growing in a micro-well plate were fixed and stained with an antibody specific for phosphorylated ERK (green). The wells were co-stained with a cell-dye (red) to monitor population size. The ratio between the two signals (green/red) marks the ERK activity normalized to cell number. **(C)** Vemurafenib impact on ERK signaling at different concentrations. The bars mark the relative pERK levels 1 h after treatment. **(D)** Vemurafenib impact on ERK signaling after multiple days of treatment. In both cell-lines, signaling recovers and rebounds to intermediate levels already within the first 24 h (reactivation period is marked by shaded area). **(E)** Vemurafenib impact on ERK signaling after multiple hours of treatment. In both cell-lines, signaling inhibition is maximized 1 h after treatment and then marginally recovers within 2 h (shaded area). In all graphs **(C–E)**, green bars mark the relative pERK levels after treatment averaged over 3 biological replicates and thin black bars mark standard deviation. The relative ERK activity is calculated by comparing its normalized level to the minimum and maximum levels observed. Stars mark pERK levels that are statistically higher than the maximum inhibition observed (one-tailed t-test with a Bonferroni corrected *p*-value for multiple statistical tests).

The recent observations of drug resistance in a rare cell population through semi-coordinated high expression of specific genes, such as growth factor receptors, agree with previous seminal work that revealed that rewiring of signaling downstream to BRAF is a key step underlying adaptive resistance (Lito et al. 2012). This study tested vemurafenib efficacy in diverse BRAF^{V600E} mutated cell-lines and showed that although vemurafenib was initially efficient in inhibiting downstream signaling, the shutdown was only transient. Within 24–48 h of drug administration, ERK phosphorylation levels rebound until they stabilized on a new intermediate steady-state level. ERK reactivation was suggested to originate from rewiring of the upstream signaling network that relieves an ERK-dependent negative feedback that suppresses ligand-dependent signaling in untreated cells. The relief of ERK-dependent negative feedback reactivates signal transduction through the generation of vemurafenib-resistant RAF dimers and culminates in renewed cell proliferation. In agreement with this mechanism of rewiring, co-

administration of a MEK inhibitor with the targeted BRAF inhibitor mitigated signaling reactivation (Lito et al. 2012). Additional observations showing that stromal cell secretion of hepatocyte growth factor can elicit resistance against RAF inhibitors further support the model of resistance through ligand-dependent signaling upon treatment (Straussman et al. 2012).

Given the widely accepted model of signaling reactivation and the discovery of a rare cell population that is drug resistant, an important open question that arises is therefore how signaling reactivation transpires at the single-cell resolution.

However, since most studies that investigated signaling rewiring relied on bulk assays, methods that pool large cell populations together for measurements, the dynamics of signaling rewiring at the single-cell resolution remain elusive. The diagram in Fig. 6.1A shows three alternative models of single-cell behaviors that are consistent with existing observations of signaling reactivation in bulk population assays. Under the *homogenous model* (Fig. 6.1A, lower-left panel) signaling rewiring is homogenous across all cells irrespective of their drug resistance state (only resistant cells later continue to proliferate). Under the alternative models, cell behavior is heterogeneous. In one heterogeneous model (Fig. 6.1A, lower-middle panel), cell-to-cell variation already emerges at the initial signaling inhibition phase. In this model signaling in resistant cells remains unaffected by the drug and a profile of signaling reactivation seemingly arises since un-inhibited cells continue to proliferate and gradually increase their proportion in the population. In a second heterogeneous model (Fig. 6.1A, lower-

right panel), all cells are initially inhibited by the drug, but resistant cells rapidly reactivate signaling.

In this study we examined how changes in ERK signaling transpire upon treatment at the single-cell resolution and tested how signaling dynamics in individual cells correlate with their ultimate proliferation capacity (drug resistance). Given observations that resistant cells highly express multiple growth factor receptors prior to drug administration (Shaffer et al. 2017), we hypothesized that rewiring of signaling will take place almost instantaneously for a small sub-population of cells (following the homogenous inhibition, heterogeneous reactivation model in Fig. 6.1A). We reasoned that this prediction can be tested by monitoring signaling dynamics within the first few hours of treatment using a live-cell signaling reporter. We further predicted that single-cell measurements will allow us to observe signaling reactivation much earlier than was previously reported in the literature since previous studies relied on pooled population measurements that “average out” heterogeneous cellular responses. Lastly, we tested if signaling dynamics immediately after treatment are good predictors of ultimate cell fate (drug resistance). We tested our predictions using a combination of bulk-population measurements and high-throughput time-lapse microscopy experiments. These approaches allowed us to monitor signaling dynamics in thousands of individual cells and rigorously quantify cell–cell heterogeneity. In agreement with our prediction, we observed that despite uniform initial inhibition, early reactivation of ERK signaling is observed within an hour for a small subpopulation of cells and that these cells indeed proliferate more than

non-recovering neighboring cells after multiple days of drug treatment. We concluded that the identification of a rapidly recovering subpopulation supports the hypothesis that a small fraction of melanoma cells pre-exists in a drug-resistant state and reconciles the widely accepted model for drug recovery through rewiring of signaling with recent observations made in single cells.

6.4 Results

6.4.1 Bulk measurement of signaling inhibition and reactivation

The effect of BRAF inhibition on MAPK signaling is typically monitored by testing the phosphorylation state of the downstream kinase ERK with semi-quantitative biochemical methods, such as Western blots, that measure the phosphorylation state averaged over large cell populations (e.g.(Lito et al. 2012; Haferkamp et al. 2013)). In order to similarly quantify ERK phosphorylation levels (pERK) in bulk populations we used the in-cell-Western (ICW) method previously used by us and others to monitor ERK signaling (Bugaj et al. 2018). Figure 6.1B shows a representative image of an ICW assay that measured ERK phosphorylation levels 1 h after adding vemurafenib. As the figure shows, we used a pERK antibody (marked in green) and a non-specific cell stain (marked in red) to simultaneously monitor ERK phosphorylation and cell confluency in fixed cell cultures. Using this method, we were able to infer the relative ERK activity, normalized to population size, across a range of drug concentrations. Figure 6.1C shows the drug response curve, as measured by pERK levels, in two melanoma cell-lines harboring the oncogenic BRAF^{V600E} mutation. A375 is a widely studied melanoma cell-line and WM983B is a cell-line that was recently derived from

a patient melanoma tumor (Shaffer et al. 2017). We found that the drug has reduced efficacy at a dose of 67.5 nM for A375 cells and 125 nM for WM983B cells (one-tailed t-test with a Bonferroni corrected p-value of 0.005 for ten statistical tests). We concluded that in our experimental system, vemurafenib is a potent signaling inhibitor at sub-micromolar concentrations for both cell-lines.

The previous discovery of signaling reactivation after BRAF^{V600E} targeted inhibition revealed that pERK levels can rebound within a few hours. However, in some cell-lines signaling reactivation can take significantly longer and transpire only after multiple days. In order to validate that signaling reactivation exists in our cells and estimate the timescale required for it to take place, we quantified pERK levels at 24 h intervals over 72 h in both cell lines. Figure 6.1D shows the results of these bulk population measurements. We observed that for both A375 and WM983B cells maximum inhibition was 1 h after the drug was added and that signaling levels considerably rebounded within 24–48 h reaching an intermediate level (one-tailed t-test with a Bonferroni corrected p-value of 0.01 for five statistical tests). The results for the A375 cell-line agree with previous observations made for this cell-line (Lito et al. 2012). Since some recovery already takes place within the first 24 h, we next decided to repeat the experiment and monitor the dynamics of pERK at a high time resolution within the first few hours of adding the drug (Fig. 6.1E). The experiments revealed that pERK levels decay rapidly after treatment and that inhibition is maximized after an hour. However, the experiments also revealed that signaling inhibition starts to slowly weaken after 2 h of treatment (one-tailed t-test with a Bonferroni corrected p-

value of 0.01 for five statistical tests). Taken together our results show that vemurafenib leads to transient signaling inhibition with initial signaling reactivation within 2 h of treatment. This observation is shorter than the previously observed time window of 4 h identified using a Western blot (Lito et al. 2012). We note however, that since our ICW approach relies on measurement of pERK level in bulk populations, it is insufficient to discriminate between the signaling rewiring models presented in Fig. 6.1A.

6.4.2 Bulk measurement of growth arrest and recovery

Cells harboring the BRAF^{V600E} mutation proliferate independently from growth-factors due to their oncogenic mutation. Previous work revealed that although targeted inhibition of the mutated BRAF initially leads to growth arrest, this arrest is transient and that growth-factor dependent proliferation can be later observed (Shaffer et al. 2017; Lito et al. 2012). We tested if transient arrest existed in the two cell-lines we used by time-lapse microscopy. We reasoned that microscopy observations will allow us to evaluate the magnitude of recovery and evaluate the minimal time interval needed for growth recovery. Towards this aim we used high-throughput time-lapse fluorescence microscopy and quantified changes in the cell numbers by automated image analysis. Quantifying growth by changes in cell number is preferable to using culture confluence since cellular morphology is noticeably influenced by vemurafenib. Figure 6.2A shows our experimental setup. We plated cells that constitutively expressed a fluorescent nuclear tag at low confluence in a micro-well plate and imaged them over a week-long experiment at 4-h intervals. To minimize potential effects of gradual decreases in drug potency

during prolonged incubation due to drug degradation or cell metabolism, we replenished half of the drug-containing media every day. This treatment also ensures that extracellular growth-factors will not be spent by the cultured cells and that dead cells will be regularly removed. Extracellular growth factors were previously shown to be important for growth recovery during vemurafenib treatment (Straussman et al. 2012). We then used automated image analysis of fluorescence images to calculate the number of individual nuclei in each imaging field. Figure 6.2B shows representative phase and fluorescence images of a single field of view and a mask image showing the nuclei detected after segmentation of the fluorescence image.

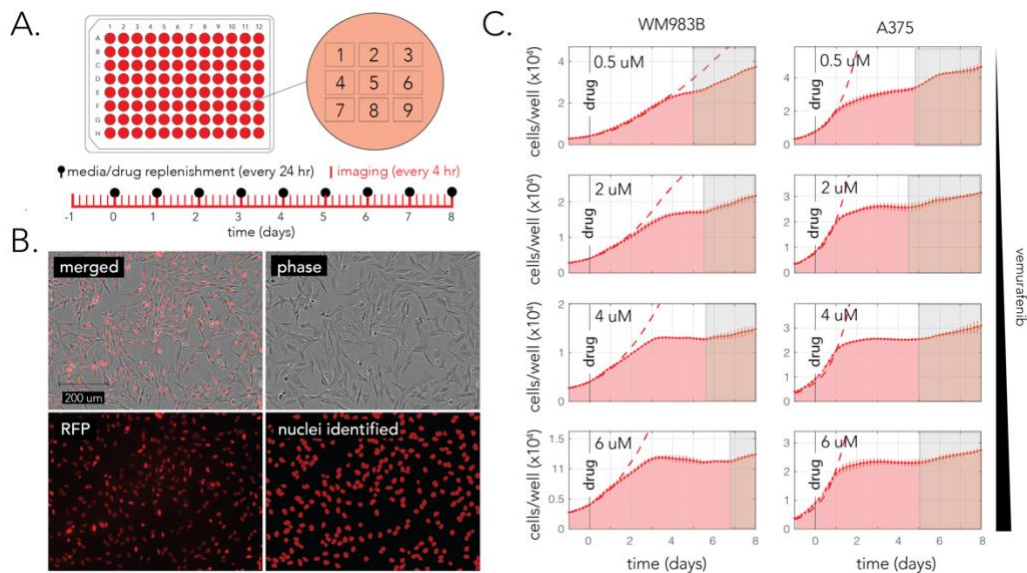


Figure 6.2 Transient growth arrest after vemurafenib treatment in bulk populations.

(A) A diagram of the drug treatment and imaging protocol. Cells were plated in a 96-well microplate and drug was replenished every 24 h. Each well was imaged with nine non-overlapping imaging fields and treatments were applied in duplicates. (B) Representative microscopy images and the paired image analysis for counting nuclei number in each imaging field. (C) Changes in cell number

over 8 days of drug treatment across multiple drug concentrations. The red area shows the inferred number of cells (per well) in drug treated wells (the bars show the standard deviation across biological replicates). The dashed red line shows the number of cells in a control (no-drug) experiment. The gray area shows the period with notable growth recovery (determined by eye). We observed growth recovery in both cell-lines, across all drug concentrations.

We monitored the growth dynamics of cells in the presence of multiple drug concentrations. We focused on a concentration range that we previously established as fully inhibitory according to pERK levels we observed 1 h after adding the drug (Fig. 6.1C). Figure 6.2C shows the results of our experiments across the four drug concentrations in the two tested cell-lines. We observed overall similar trends in growth, arrest, and recovery across all experiment conditions. We observed an almost complete growth arrest, as compared with untreated control cultures within a day or two of drug treatment (marked by a dashed line in Fig. 6.2C). However, we also detected a slow but reproducible recovery from growth arrest (marked by shaded areas in the graph). The initial response time, as measured by separation between the growth curve of treated and untreated cells, ranged from 24 h in A375 cells to 48 h in WM983B cells. This difference likely originates from the different generation time of the two cell-lines and the time it takes cells to complete the cell-cycle before arresting at the G1 stage due to the drug (Salerno et al. 2010). Importantly, we observed that in both cell-lines the arrest in growth was transient (1–5 days) and that the arrest period was inversely proportional to drug concentration. Lastly, it was also evident that at least in the experiment's time period, growth recovery was modest, with resumed growth being substantially slower than the initial growth without the drug. In agreement with previous results showing growth recovery dependence on

extracellular growth factors, we also observed a correlation between growth recovery and media replenishment frequency (Figure 6.3). Taken together, these results confirm that growth arrest is indeed transient, and establish that growth can resume very early after the arrest. Establishing this timeline was key for designing the subsequent time-lapse microscopy experiments that monitored both signaling and growth.

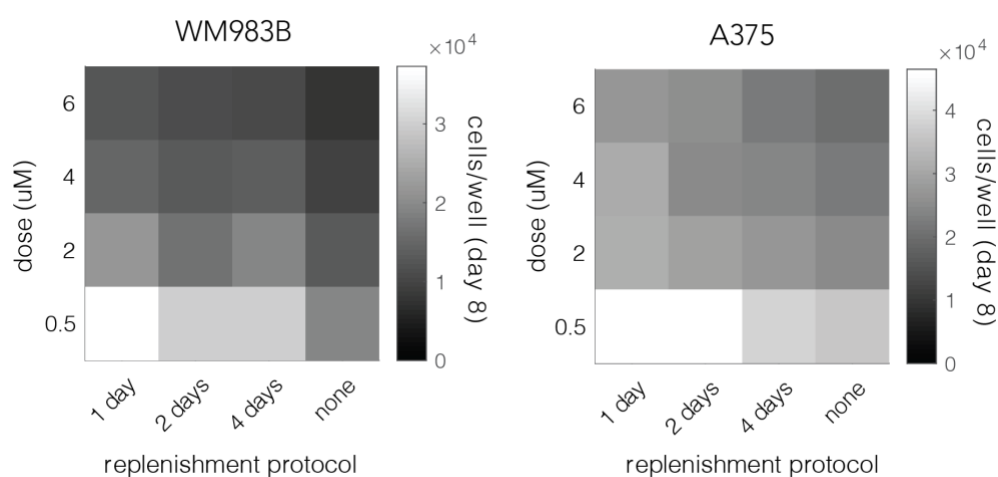


Figure 6.3 Growth recovery depends on drug concentrations and media replenishment protocol.

The heatmaps show the number of cells observed after eight days of drug treatment. In both cell-lines we observed that culture growth (drug resistance) was inversely correlated with the drug concentration and positively correlated with frequency of replenishment (half the well volume was replenished with new drug-containing media). The grayscale reflects the average cell numbers counted in two biological replicates.

6.4.3 Single-cell measurement of signaling inhibition and reactivation

Bulk measurements of signaling dynamics and cell growth allowed us to outline the relevant period for signaling reactivation and growth recovery (Figs. 6.1, 6.2). However, these experiments were insufficient for clarifying the dynamics of

individual cells and determining the underlying signaling rewiring model (Fig. 6.1A). Multiple previous studies by us and many others used translocation-based fluorescent reporters to monitor signaling activity in individual live-cells (e.g. (Shraga et al. 2019; Mitchell et al. 2015)). Such reporters alter their subcellular localization in response to a posttranscriptional modification, such as phosphorylation by the kinase of interest, and therefore provide an almost instantaneous marker of the signaling status without hampering cell viability. In order to monitor signaling downstream to BRAF we used a previously developed fluorescent kinase-translocation-reporter (KTR) that localizes primarily in the cytoplasm when ERK is active and pERK level is high (Regot et al. 2014) (Figure 6.4). In contrast, when BRAF activity is inhibited and pERK levels drop, the KTR localizes primarily in the nucleus. Figure 6.5A shows our experimental setup. We plated cells that constitutively expressed a fluorescent nuclear tag and the KTR-ERK reporter at low confluence in micro-well plates and imaged the cells every 10 min over 4 h after treatment. We decided to conduct these experiments with WM983B cells since the rare subpopulation of resistant cells was well characterized in this cell-line (Shaffer et al. 2017).

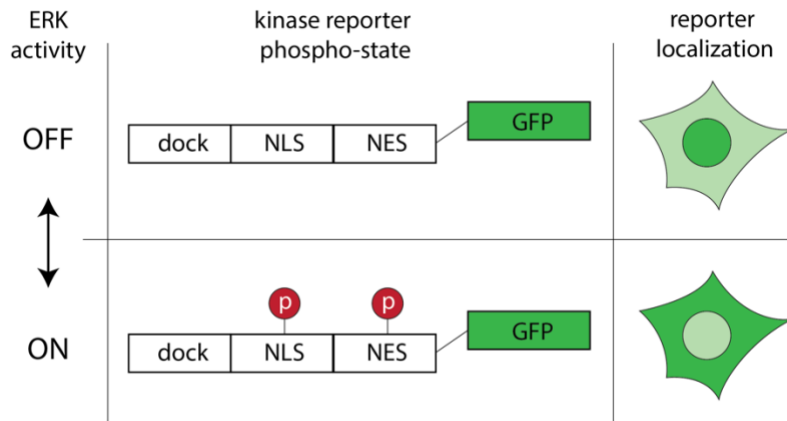


Figure 6.4 The molecular mechanism underlying the kinase translocation reporter (KTR) previously developed by the Covert lab and used in this study.

The reporter was constructed by fusing a kinase docking domain with a nuclear localization sequence (NLS), a nuclear exit sequence (NES), and a fluorescent protein (GFP). Active ERK (pERK) binds to the docking domain of the reporter and phosphorylates the NLS and NES. Phosphorylated NLS and NES favor the translocation of the reporter to the cytoplasm.

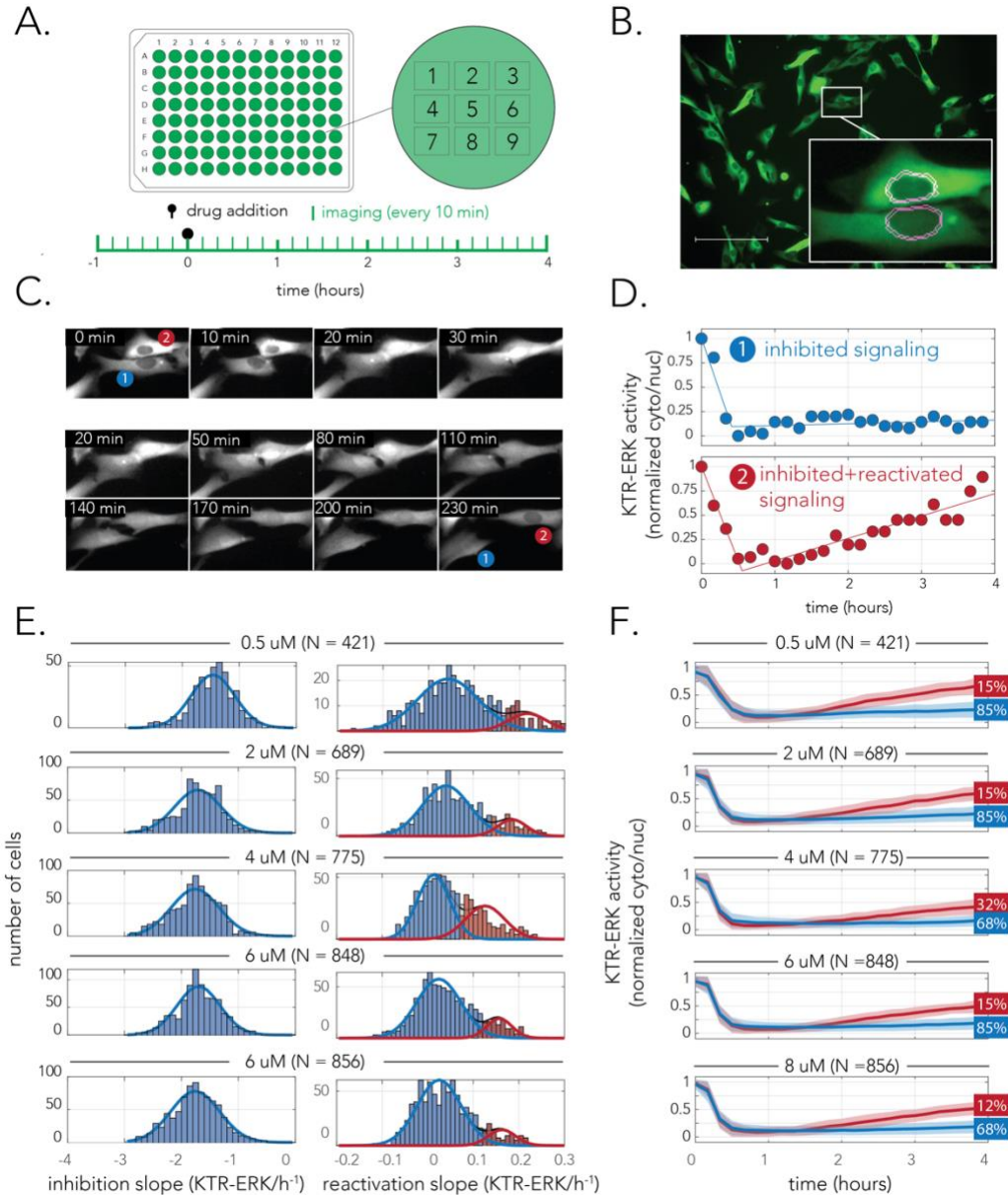


Figure 6.5 Single-cell measurements of signaling reveal homogenous inhibition yet heterogeneous signaling reactivation in WM983B cells.

(A) A diagram of the drug treatment and imaging protocol. Cells were plated in a 96-well microplate and were imaged every 10 min over 4 h after drug treatment. (B) Representative microscopy image of the translocation fluorescent reporter used to monitor ERK signaling. Image analysis was used to segment both the nucleus region of each cell and a narrow ring surrounding it (to evaluate reporter intensity in the cytoplasm). (C) Microscopy images of changes in KTR-ERK in two neighboring cells after drug treatment. Upon 20 min of treatment the fluorescent reporter translocated to the nucleus in both cells (indicating ERK signaling is inhibited). The reporter partially translocated back to the cytoplasm

after a few hours in cell #2. **(D)** Quantification of reporter localization in the two cells shown in **(C)** over time reveals different signaling reactivation in neighboring cells. The markers show the ERK activity relative to the range observed during the experiment as determined by the KTR-ERK reporter. The lines show the two-phase linear model that was fitted to the data points collected. Signaling reactivation starts as early as after 1 h in cell #2 while signaling remains inhibited in cell #1. **(E)** Cellular heterogeneity is observed in signaling reactivation but not in initial signaling inhibition. The histograms show the slopes of signaling inhibition (left panels) and signaling reactivation (right panels) across hundreds of co-cultured cells. The histograms were fitted to both unimodal and bimodal gaussians to test for heterogeneity and determine if dynamics were homogenous (unimodal) or heterogeneous (bimodal) with a statistical test. The histograms are colored with one or two colors according to the fit result and the underlying distributions are shown in thin blue and red lines. **(F)** Signaling profiles of all cells grouped and colored according to the bimodal distributions of signaling reactivation. The lines mark the average signaling status for the cell group and the shaded area shows the standard deviation. The numbers at the end of the graphs show the percentage of cells belonging to the cell group.

We developed an automated image analysis pipeline for monitoring signaling dynamics in individual cells in our time-lapse microscopy experiments. The approach relied on monitoring KTR-ERK fluorescent intensity in each of the imaged cells in addition to tracking cells over consecutive time points of the experiment (see “Methods” section). Figure 6.5B shows a representative image of cells expressing the KTR-ERK reporter and our approach for image segmentation (Fig. 6.5B inset). We segmented the nuclear region and a thin ring-shaped region around it to infer the ERK activity from the KTR-ERK intensity in the nucleus and cytoplasm (Kudo et al. 2018). Figure 6.5C shows a montage of two cells over the 4 h of the experiment as the KTR-ERK reporter translocated from the cytoplasm to the nucleus in response to BRAF inhibition. As the images show, KTR-ERK translocation can be easily detected within 20 min of adding the drug. Figure 6.5D shows the signaling dynamics we inferred with our image analysis pipeline for the cells presented in the montage. As the figure shows, although

signaling in the two cells had a similar profile in the first hour, they greatly differed in later time points—signaling clearly recovered in cell #2 while it remained fully inhibited in cell #1. In order to capture the temporal dynamics of the signaling profile we fitted the observations made in each cell to a phenomenological model with only three parameters. The signaling dynamics are fitted to a two-stage linear model that capture the slope of signaling inhibition, and the slope of signaling reactivation (or plateaued inhibition). The fitted model for the two cells is represented by the line graphs in Fig. 6.5D. It is important to note that although more complicated models can surely be used to fit the data, reduction of the dynamics to a two-phase linear model worked very well and was sufficient for our purposes.

The experimental setup we used, coupled with our automated image analysis pipeline, allowed us to simultaneously monitor the signaling dynamics of hundreds of cells in a single experiment. We therefore reasoned that this approach would allow us to identify if signaling dynamics are homogenous or heterogeneous across identically treated cells that are cultured together in a single micro-plate well. Towards this goal we decided to evaluate the regularity of signaling dynamics by inspecting the distribution of the signaling inhibition slopes and the distribution of the signaling reactivation slopes. We expected that homogenous behaviors across the cell population will transpire as unimodal distribution while heterogeneous behaviors will transpire as multimodal distribution. Figure 6.5E (and Figure 6.6A) shows the histograms of signaling inhibition and reactivation slopes in five different drug concentrations, all of

which are expected to be fully inhibitory by their initial impact (Fig. 6.1C). We used a Kolmogorov–Smirnov statistical test to see if the histograms better fit a single unimodal gaussian or a bimodal gaussian (mixture of two unimodal gaussians). This analysis revealed that the initial stage of drug response, signaling inhibition, is homogenous in the cell population across the entire drug concentration range. However, the same analysis revealed that signaling reactivation is heterogeneous and fits a mixture of two unimodal gaussian distributions (marked in red and blue lines in the figure). This bimodal distribution indicated that the population consists of two differently behaving subpopulations: signaling remains inhibited for the majority of cells yet it gets reactivated in a minority of them. The signaling dynamics of these two subpopulations is similar to that presented for the two cells in Fig. 6.5D. Importantly, our analysis also allows us to estimate the proportion of each subpopulation from all imaged cells (marked by the percentage numbers at the edge of the graphs in Fig. 6.5F and Fig. 6.6B). These estimations showed that the proportion of the recovering subpopulation is almost identical across all tested drug concentrations (averaging around 15%).

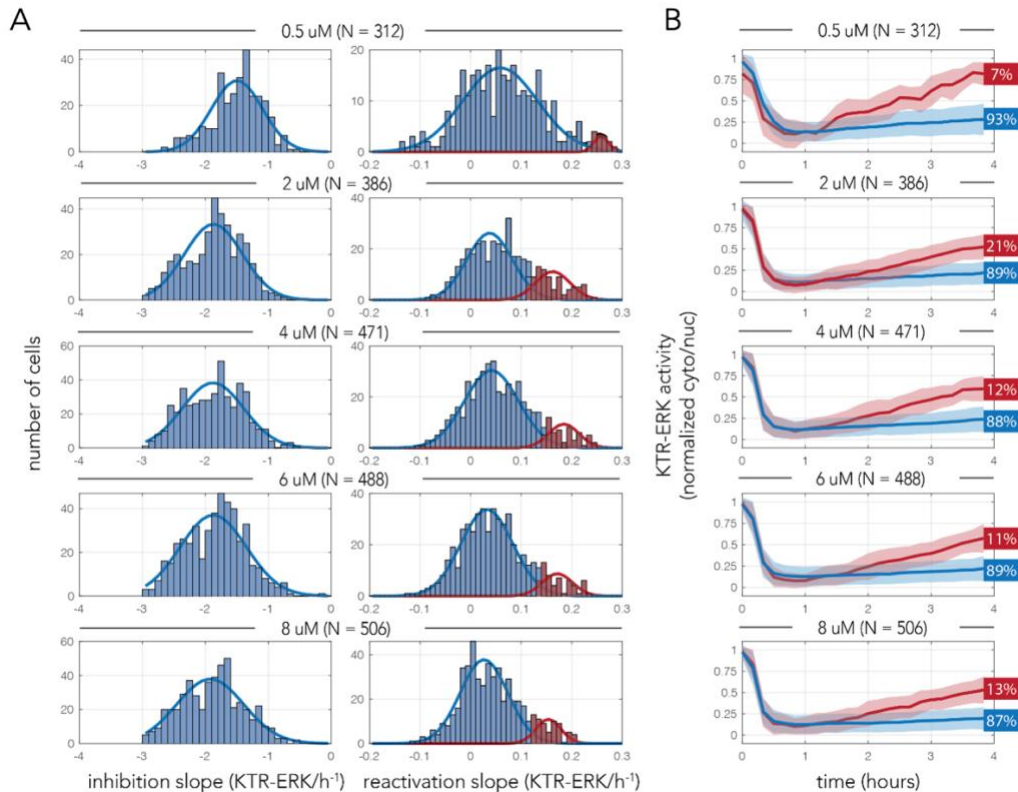


Figure 6.6 Biological replication of the experiment presented in Figure 6.5.

(A) Cellular heterogeneity is observed in signaling reactivation but not in initial signaling inhibition. The histograms show the slopes of signaling inhibition (left panels) and signaling reactivation (right panels) across hundreds of co-cultured cells. The histograms were fitted to both unimodal and bimodal gaussians to test for heterogeneity and determine if dynamics were homogenous (unimodal) or heterogeneous (bimodal) with a statistical test. The histograms are colored with one or two colors according to the fit result and the underlying distributions are shown in thin blue and red lines. (B) Signaling profiles of all cells grouped and colored according to the bimodal distributions of signaling reactivation. The lines mark the average signaling status for the cell group and the shaded area shows the standard deviation. The numbers at the end of the graphs show the percentage of cells belonging to the cell group.

Taken together, our single-cell results indicate that WM983B cells can be stratified into two subpopulations by their signaling response to vemurafenib within the first few hours of treatment—a large group of cells that remains inhibited and a small group capable of reactivating signaling after initial

inhibition within an hour of treatment. These dynamics support the homogenous inhibition and heterogeneous reactivation model of signaling rewiring presented in Fig. 6.1A. Furthermore, the short time-scale in which the differences between the two subpopulations manifests itself is too short to be explained by a different transcriptional program that is triggered post-treatment. The differences in signaling rewiring therefore likely reflect dissimilar transcriptional states that pre-existed the drug. This hypothesis is further supported by the observation that the same fraction of cells recovered irrespectively from the drug concentration used.

6.4.4 Single-cell measurement of growth recovery

Our long-term measurement of bulk population growth after drug treatment revealed that the cell cultures are transiently arrested and then resume growth at a slow rate (Fig. 6.2C). Given the cell–cell heterogeneity we observed in signaling reactivation, a key question that arises is whether the signaling profile of a cell within hours of drug treatment is a good predictor of its proliferation capacity. To address this question, we combined the experimental approaches we used previously (Figs. 6.2, 6.5) in a two-phase time-lapse microscopy experiment. In this experiment we first monitored early signaling dynamics, at a high time resolution, and then continued and monitored the proliferation of individual cells over 7 days at a lower time resolution (see “Methods” section). Since accurately tracking individual cell lineages over multiple days was impractical by automated image analysis (due to cell movement, division, and detachment) we tracked lineages by eye after observing time-lapse microscopy images.

In these experiments we focused on the WM983B cells treated with a single inhibitory concentration of vemurafenib throughout the experiment (6 uM) and focused on a representative sample of 180 cells chosen by their early signaling dynamics. Figure 6.7A shows the signaling dynamics of the subset of cells that we chose for this analysis (taken from two independent experiments). A sample group of 120 cells was chosen to represent the major subpopulation of non-recovering cells (blue lines in Fig. 6.7A). This group was further subdivided into a group of 60 fully inhibited cells (light blue lines) and a group of 60 inhibited cells with a marginal recovery (dark blue lines). A sample group of 60 cells was chosen to represent the minor subpopulation of signaling reactivating cells (red lines in Fig. 6.7A). We next counted how many cells descended from the 180 monitored cells during a period of seven days of drug treatment (a period sufficient to detect growth recovery, Fig. 6.2C). Figure 6.7B shows bar plots representing the lineage size originating from each tracked cell after sorting the cells by their number of decedents and coloring the bars according to the early signaling dynamics. As the figure shows, we observed that cells characterized by early signaling recovery (red bars) had significantly larger lineages than non-recovering cells (blue bars). A non-parametric statistical test on the rank order of cells from the two groups rejected the null hypothesis that they are characterized by identical lineage sizes (p -value $< 10^{-18}$, Wilcoxon Rank Sum Test). The significant correlation between early signaling dynamics and lineage size can also be visualized by comparing the average number of decedents of cells from each group (Fig. 6.7C, 6.8), with recovering cells giving rise to an average lineage size of 6.6 cells within a week of

drug treatment and non-recovering cells giving rise to an average 2.7 cells in the same period. A statistical test considering the sample groups rejected the hypothesis that the two groups have similar lineage sizes (p-value $< 10^{-20}$, two-tailed t-test). It is important to note the fact that even non-recovering cells are not completely arrested (lineage size larger than one) is compatible with the observation that growth arrest is not immediate with drug treatment (Fig. 6.2C) and that growth arrest due to vemurafenib takes place at the G1 stage (Salerno et al. 2010). Taken together the results of this experiment establish a direct association between the early reactivation of signaling and ultimately cell proliferation capacity. These direct observations of individual cells throughout the treatment period strongly supports the heterogeneous model of drug recovery presented in Fig. 6.1A.

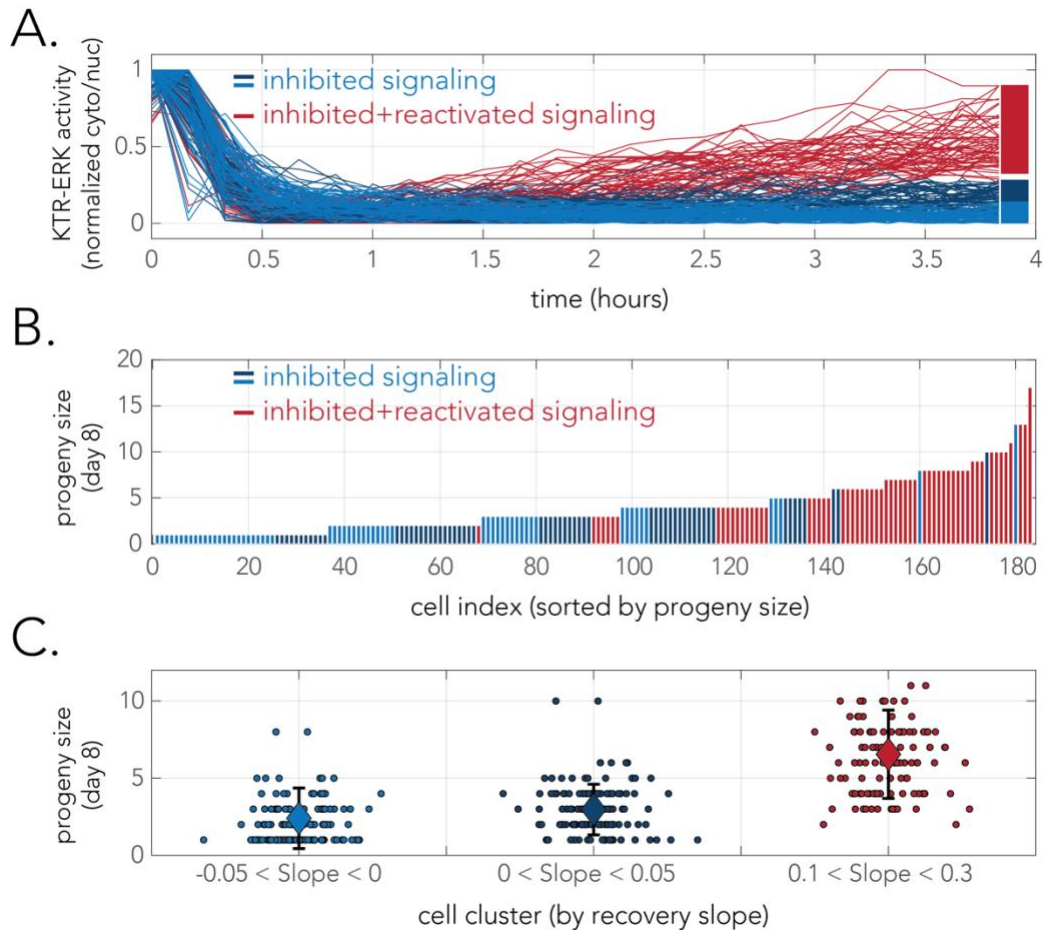


Figure 6.7 Single-cell measurements connect early signaling recovery with long term drug resistance.

(A) Signaling dynamics at the first 4 h after drug treatment of 180 cells treated with inhibitory concentration of vemurafenib (6 μ M). The colors mark the signaling recovery capacity of the cells (recovering cells in red and non-recovering cells in blue). The identity of the cells was determined solely according to the slope of the signaling curve after the initial inhibition phase. The non-recovering cells were further subdivided to strongly inhibited cells (light blue) and weakly inhibited cells (light blue). (B) Lineage sizes descending from the 180 cells after seven days of drug treatment. The bars show the lineage descending from single cells and the bar colors mark initial signaling recovery of the ancestor cell. (C) Lineage sizes descending from ancestor cells according to the signaling recovery group. Each point marks the lineage descending from a single ancestor cell and the large diamond markers show the average lineage size. Error bars represent standard deviation.

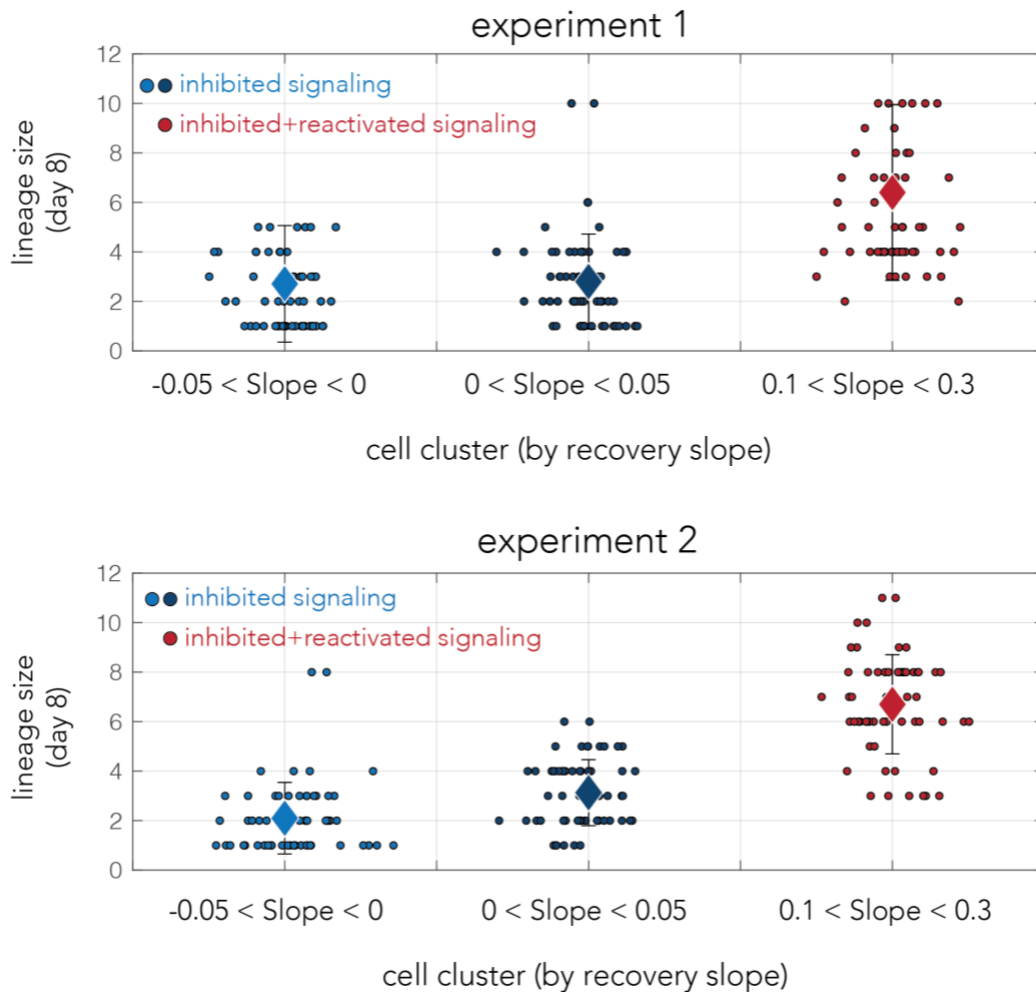


Figure 6.8 Biological replicates used in the experiment presented in Figure 6.7.

The top and lower panels show the observations made in two separate experiments (performed on different days) that measured the correlation between early dynamics of ERK activity and the long-term proliferation of individual cells. The y-axis shows the lineage sizes descending from ancestor cells according to the signaling recovery group. Each point marks the lineage descending from a single ancestor cell and the large diamond markers show the average lineage size. Error bars represent standard deviation.

6.5 Discussion

Phenotypic differences between cells are a defining characteristic in cancer that greatly increases disease complexity and challenges its treatment. Numerous studies over the past decades established that cell-to-cell heterogeneity can engage key oncogenic pathways underlying disease and can therefore pose significant challenges for personalized cancer medicine (Burrell et al. 2013). While much attention is currently dedicated to uncovering phenotypic diversity that arises due to genetic differences between cells in the same tumor, recent works uncovered that epigenetic mechanisms, operating differently within isogenic cells, also play a significant role in promoting heterogeneity with clinical implications (Black and McGranahan 2021). Here we focused on a widely used in-vitro model for studying adaptive drug resistance—growth recovery after treatment with a targeted therapy (vemurafenib) against a common oncogenic BRAF mutation (Lito et al. 2012). We chose this system since previous work indicated that it features highly heterogeneous drug resistance and pointed to a plausible underlying epigenetic mechanism—differences between isogenic cells in the expression of key genes in the cellular network responding to growth factors (Shaffer et al. 2017). Although such transcriptional differences are potentially compatible with the widely accepted model of vemurafenib resistance through post-treatment reliance on extracellular growth signals (Lito et al. 2012; Straussman et al. 2012), direct evidence of this connection is currently missing. Specifically, there are no observations of heterogeneity in signaling reactivation at the single-cell level and no evidence of an association between the profile of

signaling of individual cells and their proliferation capacity during drug treatment (Mitchell et al. 2015; Karanam et al. 2012; Purvis et al. 2012; Spencer et al. 2009; Lee et al. 2014).

It is important to note that there is a considerable challenge in gathering such direct evidence of a connection between cellular processes that manifest on such different time scales that might also be highly heterogeneous across the population. First, work on heterogeneity requires single-cell assays that are typically more demanding than bulk population measurements and typically rely on “end-point” assays that require fixing the cells (e.g., immunofluorescence) or at least significantly perturbing them (e.g., live-cell cytometry). Due to this limitation, multiple independent observations, as signaling and subsequent proliferation, can rarely be done on the same exact cell. Additionally, in this specific model system, there is the need to simultaneously monitor multiple cellular processes, from cell-to-cell differences in the transcriptional program that exist for days before the drug treatment (Shaffer et al. 2017), followed by rewiring of the signaling network that can transpire within minutes of drug treatment, and eventually with changes in cell proliferation that take multiple days (Lito et al. 2012). Indeed, this challenge of bridging events that transpire on multiple time-scales is common to many model systems in cell biology (Tsabar et al. 2021). Here we addressed this challenge by performing time-lapse microscopy experiments, at different time resolutions, in a week-long experiment while maintaining ideal growth conditions (without any perturbations to the treated cells). This methodology has proven valuable in bridging such time gaps between

a fast signaling event, transcriptional response and cell fate decisions in the past (e.g. (Mitchell et al. 2015; Karanam et al. 2012; Purvis et al. 2012; Spencer et al. 2009; Lee et al. 2014). While being demanding, this approach allowed us to make single-cell observations and gather direct evidence establishing a clear connection between early signaling dynamics and long-term proliferation capacity. These observations fill in a critical piece of the puzzle that connected the widely accepted rewiring model, suggested a decade ago, to population heterogeneity in drug sensitivity that was reported only very recently (Shaffer et al. 2017).

The experimental approach we present here can be further used to address additional open questions in this important model system. For example, while we established a connection between signaling dynamics and drug resistance, additional work will be required to establish a connection between the transcriptional state of cells pre-treatment and their signaling dynamics post-treatment. However, this connection may not be easy to establish given that cells existing in the resistant state prior to the drug may themselves be highly heterogeneous, with different resistant cells over-expressing a different set of resistance genes (Shaffer et al. 2017). Another important question that might be addressed using our microscopy approach concerns the stability and inheritance of the resistance state. For example, it is possible to monitor if kin-cells that descended from a common ancestor prior to treatment share a common signaling profile as a function of the time that elapsed from cell division to drug treatment. Such measurements can help refine parameters of the state model previously developed for these melanoma cells (Shaffer et al. 2017). Alternatively,

information regarding other signaling pathways can also be collected using orthogonal live-cell reporters in order to provide direct evidence of additional early signaling events that predict long-term drug resistance. Such investigation is highly interesting in the context of the JNK pathway that was previously suggested to be implicated in vemurafenib drug resistance (Fallahi-Sichani et al. 2017, 2015).

6.6 Methods

6.6.1 Cell-lines and media and antibodies

A375 (ATCC) and WM983B (Arjun Raj lab) cells were cultured in DMEM supplemented with 5% fetal bovine serum (Gibco FBS cat# 26140-079) and incubated at 37 °C with 5% CO₂. All experiments were conducted by plating cells on 96-well plates 12–18 h before starting the experiment. BRAF targeted inhibition was achieved with Vemurafenib (Selleckchem, PLX4032). Time lapse microscopy experiments were performed with cells constitutively expressing fluorescent reporters after infection with lentivirus.

6.6.2 Live-cell fluorescent reporters

Plasmid constructs for mammalian expression were cloned into the pHR lentiviral backbone similarly to our previous work (Shraga et al. 2019). We cloned the mammalian histone H2B-mRuby and the pathway reporter KTR(ERK)-mEGFP (Regot et al. 2014) into separate plasmids. Lentivirus was produced by co-transfecting the pHR plasmids and vectors encoding packaging proteins (pMD2.G and p8.91) using the Fugene 6 HD transfection reagent in HEK-293T cells plated

in 6-well plates at ~ 70% confluency. Viral supernatants were collected 2 days after transfection and 0.45 μm filtered and used for transduction immediately. A375 and WM983B cells were cultured in 5% fetal bovine serum in DMEM at 37 °C with 5% CO₂ in a humidified incubator. For viral transduction, cells were plated in 6-well plates to achieve ~ 20% confluency at the time of infection. For lentiviral transduction, we added 50 μL of virus supernatant directly to cells. Viral media was replaced with growth media 24 h post infection.

6.6.3 72-h dose response (ICW)

A375 and WM983B were seeded in 96-well plates (Greiner, 655090; 7500 cells/well) in 100 μL media the day before treatment. Vemurafenib (Selleckchem, PLX4032; 2 \times in 100 μL media) was added to wells in triplicate with doses 0 μM , 0.031 μM , 0.062 μM , 0.125 μM , 0.25 μM , 0.5 μM , 1 μM , 2 μM , 4 μM , 8 μM for 1 h. Vemurafenib media was removed and wells were washed 3 times with fresh media before replenishing wells with 200 μL media. After 72 h, cells were analyzed with an in-cell western.

6.6.4 pERK peak inhibition and recovery (ICW)

A375 and WM983B were seeded in 96-well plates (Greiner, 655090; 10,000 cells/well) in 100 μL media in triplicate the day before treatment. Vemurafenib (Selleckchem, PLX4032; 2 \times in 100 μL media) was added for a final concentration of 0.5 μM . Cells were treated for 0, 0.25, 0.5, 1, 2, and 4 h (peak inhibition and short-term recovery) and for 0, 1, 24, 48, and 72 h (long-term recovery) before analyzing with an in-cell western.

6.6.5 In-cell western

Cells were grown and treated in 96-well plates. Treated cells were fixed with 3.7% formaldehyde for 20 min at room temperature (RT) and washed 2 times with 200 uL PBS (RT). Cells were permeabilized with 200 uL/well 0.5% Triton X-100 (10 min, RT) and blocked with 150 uL/well Odyssey Blocking Buffer (LI-COR) for 1.5 h at RT with 300 rpm shaking. Cells were incubated overnight at 4 °C with anti-phospho-Erk1/2 (Cell Signaling #4370; 1:400). Primary antibody solution was removed and wells were washed 5 times with 200 uL PBS-T (5 min each, RT, 300 rpm). A near-infrared (800 nm) fluorescent secondary antibody (LI-COR 926-32210; 1:800) and near-infrared (700 nm) cell counterstain (LI-COR 926-41090; 1:2000) were incubated with cells (1 h, RT, 300 rpm).

Secondary antibody was removed before washing wells 2 times with 200 uL PBS-T (5 min each, RT, 300 rpm). 96-well plates were imaged with LI-COR Odyssey. Signal (800 nm and 700 nm) per well was quantified using Image Studio Lite (LI-COR). Relative phospho-Erk1/2 signal per well was determined by dividing the phospho-Erk1/2 signal (800 nm) by the CellTag signal (700 nm) and replicates averaged (mean).

6.6.6 Time-lapse microscopy and automated image analysis

For imaging, cells were plated in 96-well TC-treated plates (Eppendorf) at a concentration of 3000 cells/well and were allowed to adhere over night before imaging started (12–18 h). All microscopy images were obtained using an IncuCyte S3 microscope platform (Sartorius) that is placed inside a humidified incubator with temperature and CO₂ control. We used built-in filter sets in the

IncuCyte microscope (Green/Red 4616 optical module) to monitor GFP (300 ms acquisition time), RFP (400 ms acquisition time) and phase channels. For experiments monitoring cellular signaling we imaged nine none-overlapping fields in each well with a $\times 20$ magnification objective with time intervals of 10 min. For experiments monitoring cell growth, we imaged nine none-overlapping fields in each well with a $\times 20$ magnification objective with time intervals of 1 h. All experiments were conducted with biological replicates in separate wells (indicated in the “Results” section).

Image analysis was performed similarly to our previous work (Shraga et al. 2019) with the propriety software of the IncuCyte microscope and additional custom written computer scripts developed in MATLAB (Mathworks).

Calculation of nuclei number in each imaging field was based on segmentation of the mRuby fluorescence signal with the IncuCyte software. We used Top-Hat algorithm to correct for uneven illumination. We used area and eccentricity filters to further remove fluorescent objects that are not live cells (specs of dust, cellular debris). The number of nuclei per well was calculated by summing the number of objects detected over all imaging fields and multiplying a constant conversion coefficient (the coefficient is the ratio of well surface area over the total area of the imaging fields).

To analyze signaling dynamics we first segmented the nuclei according to the mRuby fluorescence signal using the IncuCyte software and produced black/white mask images that correspond to the detected nuclei. These mask images were first

used for registration of fluorescence images (aligning consecutive images to account for systematic drift) and then used for segmenting individual cells and inferring the signaling activities (inferring the localization of the nucleus region and cytoplasm ring region and in the signal intensity (Shraga et al. 2019; Kudo et al. 2018). For image registration we used the Descriptor based registration plugin (Preibisch et al. 2010) in the Fiji distribution (Schindelin et al. 2012) of ImageJ (Schneider et al. 2012). Registered images were then used as input to a Matlab script that segmented all the objects (nuclei) in each mask image and a narrow ring region around the segmented nuclei. We subtracted the background fluorescence from GFP images and measured the 80th percentile nuclear intensity and the corresponding cytoplasm ring region. Pathway activity was calculated as the ratio between the signals of the cytoplasm ring and the nuclear region (Shraga et al. 2019). Lastly, in order to track temporal changes in signaling activity, we reconstructed continuous pathway activity for each cell over the entire time-lapse period (4 h) by tracking individual cells over time. Cell tracking was performed with a Matlab script that assigned to each nucleus the spatially closest nucleus over consecutive time points while filtering out nuclei that move to far or underwent significant changes in their area (typically corresponding to cells undergoing mitosis). By assigning highly conservative filters for removing suspicious nuclei we were able to maintain highly reliable tracks for over half of all detected nuclei (estimated by inspection of tracks for individual random cells). Nuclei that were filtered out were ones that migrated out of the imaging area, underwent mitosis or lost adherence during the experiment.

Bibliography

- Abraham G, Yuchen G, Jennifer L, Daniela P, Amanda X, S CC, S BD, Mihaela P, L SS. 2023. Major data analysis errors invalidate cancer microbiome findings. *mBio* **14**: e01607-23. <https://doi.org/10.1128/mbio.01607-23>.
- Abram K, Udaondo Z, Bleker C, Wanchai V, Wassenaar TM, Robeson MS, Ussery DW. 2021. Mash-based analyses of Escherichia coli genomes reveal 14 distinct phylogroups. *Commun Biol* **4**: 117. <https://doi.org/10.1038/s42003-020-01626-5>.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**: 415–421. <https://doi.org/10.1038/nature12477>.
- Alnahas S, Hagner S, Raifer H, Kilic A, Gasteiger G, Mutters R, Hellhund A, Prinz I, Pinkenburg O, Visekruna A, et al. 2017. IL-17 and TNF- α Are Key Mediators of Moraxella catarrhalis Triggered Exacerbation of Allergic Airway Inflammation. *Front Immunol* **8**. <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2017.01562>.
- Altshuller M, He X, MacKrell EJ, Wernke KM, Wong JWH, Sellés-Baiget S, Wang T-Y, Chou T-F, Duxin JP, Balskus EP, et al. 2024. The Fanconi anemia pathway repairs colibactin-induced DNA interstrand cross-links. *bioRxiv* 2024.01.30.576698. <http://biorxiv.org/content/early/2024/01/31/2024.01.30.576698.abstract>.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al. 2011. Enterotypes of the human gut microbiome. *Nature* **473**: 174–180. <https://doi.org/10.1038/nature09944>.
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006.0008. <https://doi.org/10.1038/msb4100050>.
- Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-Bacterial Mutualism in the Human Intestine. *Science (1979)* **307**: 1915–1920. <https://doi.org/10.1126/science.1104816>.
- Bailey TL. 2021. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**: 2834–2840. <https://doi.org/10.1093/bioinformatics/btab203>.

- Banerjee S, Tian T, Wei Z, Shih N, Feldman MD, Peck KN, DeMichele AM, Alwine JC, Robertson ES. 2018. Distinct Microbial Signatures Associated With Different Breast Cancer Types. *Front Microbiol* **9**. <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2018.00951>.
- Barnes DE. 2001. Non-homologous end joining as a mechanism of DNA repair. *Current Biology* **11**: R455–R457. [https://doi.org/10.1016/S0960-9822\(01\)00279-2](https://doi.org/10.1016/S0960-9822(01)00279-2).
- Bellacosa A. 2001. Functional interactions and signaling properties of mammalian DNA mismatch repair proteins. *Cell Death Differ* **8**: 1076–1092. <https://doi.org/10.1038/sj.cdd.4400948>.
- Bernelot-Moens C, Glickman BW, Gordon AJE. 1990. Induction of specific frameshift and base substitution events by benzo[a]pyrene diol epoxide in excision-repair-deficient Escherichia coli. *Carcinogenesis* **11**: 781–785. <https://doi.org/10.1093/carcin/11.5.781>.
- Bezine E, Malaisé Y, Loeuillet A, Chevalier M, Boutet-Robinet E, Salles B, Mirey G, Vignard J. 2016. Cell resistance to the Cytotoxic Distending Toxin involves an association of DNA repair mechanisms. *Sci Rep* **6**: 36022. <https://doi.org/10.1038/srep36022>.
- Black JRM, McGranahan N. 2021. Genetic and non-genetic clonal diversity in cancer evolution. *Nat Rev Cancer* **21**: 379–392. <https://doi.org/10.1038/s41568-021-00336-2>.
- Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, Lazarev MG, Ellis B, Carroll KC, Albesiano E, Wick EC, et al. 2015. The Bacteroides fragilis Toxin Gene Is Prevalent in the Colon Mucosa of Colorectal Cancer Patients. *Clinical Infectious Diseases* **60**: 208–215. <https://doi.org/10.1093/cid/ciu787>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P. 2012. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* **11**: 873–886. <https://doi.org/10.1038/nrd3847>.
- Boot A, Huang MN, Ng AWT, Ho S-C, Lim JQ, Kawakami Y, Chayama K, Teh BT, Nakagawa H, Rozen SG. 2018. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res* **28**: 654–665.

- Booth SC, Smith WPJ, Foster KR. 2023. The evolution of short- and long-range weapons for bacterial competition. *Nat Ecol Evol* **7**: 2080–2091. <https://doi.org/10.1038/s41559-023-02234-2>.
- Bossuet N, Guyonnet C, Chagneau C V, Tang-Fichaux M, Penary M, Loubet D, Branchu P, Oswald E, Nougayrede J-P. 2023. Oxygen concentration modulates colibactin production. *Gut Microbes* **15**: 2222437. <https://doi.org/10.1080/19490976.2023.2222437>.
- Bossuet-Greif N, Dubois D, Petit C, Tronnet S, Martin P, Bonnet R, Oswald E, Nougayrède J-P. 2016. Escherichia coli ClbS is a colibactin resistance protein. *Mol Microbiol* **99**: 897–908. <https://doi.org/10.1111/mmi.13272>.
- Bossuet-Greif N, Vignard J, Taieb F, Mirey G, Dubois D, Petit C, Oswald E, Nougayrède J-P. 2018. The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *mBio* **9**: e02393-17. <https://pubmed.ncbi.nlm.nih.gov/29559578>.
- Brash DE. 2015. UV Signature Mutations. *Photochem Photobiol* **91**: 15–26. <https://doi.org/10.1111/php.12377>.
- Brotherton CA, Balskus EP. 2013. A Prodrug Resistance Mechanism Is Involved in Colibactin Biosynthesis and Cytotoxicity. *J Am Chem Soc* **135**: 3359–3362. <https://doi.org/10.1021/ja312154m>.
- Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, Pezet D, Bonnet R. 2013. High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One* **8**: e56964. <https://doi.org/10.1371/journal.pone.0056964>.
- Buermeyer AB, Deschênes SM, Baker SM, Liskay RM. 1999. Mammalian DNA Mismatch Repair. *Annu Rev Genet* **33**: 533–564.
- Bugaj LJ, Sabnis AJ, Mitchell A, Garbarino JE, Toettcher JE, Bivona TG, Lim WA. 2018. Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. *Science (1979)* **361**: eaao3048. <https://doi.org/10.1126/science.aao3048>.
- Burnouf D, Duane M, Fuchs RP. 1987. Spectrum of cisplatin-induced mutations in Escherichia coli. *Proceedings of the National Academy of Sciences* **84**: 3758–3762. <https://doi.org/10.1073/pnas.84.11.3758>.
- Burrell RA, McGranahan N, Bartek J, Swanton C. 2013. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**: 338–345. <https://doi.org/10.1038/nature12625>.

- Camilla U, B GG, Muriel B, Leslie S, Mark T, Gregor R. 2016. The Microbiota of Breast Tissue and Its Association with Breast Cancer. *Appl Environ Microbiol* **82**: 5039–5048. <https://doi.org/10.1128/AEM.01235-16>.
- Cañas M-A, Giménez R, Fábrega M-J, Toloza L, Baldomà L, Badia J. 2016. Outer Membrane Vesicles from the Probiotic *Escherichia coli* Nissle 1917 and the Commensal ECOR12 Enter Intestinal Epithelial Cells via Clathrin-Dependent Endocytosis and Elicit Differential Effects on DNA Damage. *PLoS One* **11**: e0160374. <https://doi.org/10.1371/journal.pone.0160374>.
- Cantwell-Dorris ER, O’Leary JJ, Sheils OM. 2011. BRAFV600E: Implications for Carcinogenesis and Molecular Therapy. *Mol Cancer Ther* **10**: 385–394. <https://doi.org/10.1158/1535-7163.MCT-10-0799>.
- Cardona F, Andrés-Lacueva C, Tulipani S, Tinahones FJ, Queipo-Ortuño MI. 2013. Benefits of polyphenols on gut microbiota and implications in human health. *J Nutr Biochem* **24**: 1415–1422. <https://www.sciencedirect.com/science/article/pii/S0955286313000946>.
- Carim S, Azadeh AL, Kazakov AE, Price MN, Walian PJ, Lui LM, Nielsen TN, Chakraborty R, Deutschbauer AM, Mutalik VK, et al. 2021. Systematic discovery of pseudomonad genetic factors involved in sensitivity to tailocins. *ISME J* **15**: 2289–2305. <https://doi.org/10.1038/s41396-021-00921-1>.
- Chagneau C V., Garcie C, Bossuet-Greif N, Tronnet S, Brachmann AO, Piel J, Nougayrède J-P, Martin P, Oswald E. 2019. The Polyamine Spermidine Modulates the Production of the Bacterial Genotoxin Colibactin. *mSphere* **4**: 10.1128/msphere.00414-19. <https://doi.org/10.1128/msphere.00414-19>.
- Chang HHY, Pannunzio NR, Adachi N, Lieber MR. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**: 495–506. <https://doi.org/10.1038/nrm.2017.48>.
- Chatterjee N, Walker GC. 2017. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**: 235–263. <https://doi.org/10.1002/em.22087>.
- Chayot R, Montagne B, Mazel D, Ricchetti M. 2010. An end-joining repair mechanism in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **107**: 2141–2146. <https://doi.org/10.1073/pnas.0906355107>.
- Chen B, Ramazzotti D, Heide T, Spiteri I, Fernandez-Mateos J, James C, Magnani L, Graham TA, Sottoriva A. 2023. Contribution of pks+ *E. coli* mutations to colorectal carcinogenesis. *Nat Commun* **14**: 7827. <https://doi.org/10.1038/s41467-023-43329-5>.

- Chen J, Byun H, Liu R, Jung I-J, Pu Q, Zhu CY, Tanchoco E, Alavi S, Degnan PH, Ma AT, et al. 2022. A commensal-encoded genotoxin drives restriction of *Vibrio cholerae* colonization and host gut microbiome remodeling. *Proceedings of the National Academy of Sciences* **119**: e2121180119. <https://doi.org/10.1073/pnas.2121180119>.
- Chen J, Potlapalli R, Quan H, Chen L, Xie Y, Pouriye S, Sakib N, Liu L, Xie Y. 2024. Exploring DNA Damage and Repair Mechanisms: A Review with Computational Insights. *BioTech* **13**: 3.
- Cheng WT, Kantilal HK, Davamani F. 2020. The Mechanism of *Bacteroides fragilis* Toxin Contributes to Colon Cancer Formation. *Malaysian Journal of Medical Sciences* **27**: 9–21.
- Cole RS. 1973. Repair of DNA Containing Interstrand Crosslinks in *Escherichia coli*: Sequential Excision and Recombination. *Proceedings of the National Academy of Sciences* **70**: 1064–1068. <https://doi.org/10.1073/pnas.70.4.1064>.
- Collier-Hyams LS, Zeng H, Sun J, Tomlinson AD, Bao ZQ, Chen H, Madara JL, Orth K, Neish AS. 2002. Cutting Edge: Salmonella AvrA Effector Inhibits the Key Proinflammatory, Anti-Apoptotic NF- κ B Pathway¹. *The Journal of Immunology* **169**: 2846–2850. <https://doi.org/10.4049/jimmunol.169.6.2846>.
- Cornforth DM, Foster KR. 2013. Competition sensing: the social side of bacterial stress responses. *Nat Rev Microbiol* **11**: 285–293. <https://doi.org/10.1038/nrmicro2977>.
- Cortes-Bratti X, Frisan T, Thelestam M. 2001. The cytolethal distending toxins induce DNA damage and cell cycle arrest. *Toxicon* **39**: 1729–1736. <https://www.sciencedirect.com/science/article/pii/S0041010101001593>.
- Cougnoux A, Dalmaso G, Martinez R, Buc E, Delmas J, Gibold L, Sauvanet P, Darcha C, Déchelotte P, Bonnet M, et al. 2014a. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **63**: 1932 LP – 1942. <http://gut.bmj.com/content/63/12/1932.abstract>.
- Cougnoux A, Dalmaso G, Martinez R, Buc E, Delmas J, Gibold L, Sauvanet P, Darcha C, Déchelotte P, Bonnet M, et al. 2014b. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **63**: 1932. <http://gut.bmj.com/content/63/12/1932.abstract>.
- Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrède J-P. 2010. *Escherichia coli* induces DNA damage in vivo and triggers genomic

- instability in mammalian cells. *Proceedings of the National Academy of Sciences* **107**: 11537–11542. <https://doi.org/10.1073/pnas.1001261107>.
- Deans AJ, West SC. 2011. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* **11**: 467–480.
- Deatherage DE, Barrick JE. 2014. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq. In *Engineering and Analyzing Multicellular Systems: Methods and Protocols* (eds. L. Sun and W. Shou), pp. 165–188, Springer New York, New York, NY https://doi.org/10.1007/978-1-4939-0554-6_12.
- Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu X, DeStefano Shields CE, Hechenbleikner EM, Huso DL, et al. 2018. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science (1979)* **359**: 592 LP – 597. <http://science.sciencemag.org/content/359/6375/592.abstract>.
- Dougherty MW, Jobin C. 2021. Shining a Light on Colibactin Biology. *Toxins (Basel)* **13**: 346. <https://pubmed.ncbi.nlm.nih.gov/34065799>.
- Dougherty MW, Valdés-Mas R, Wernke KM, Gharaibeh RZ, Yang Y, Brant JO, Riva A, Muehlbauer M, Elinav E, Puschhof J, et al. 2023. The microbial genotoxin colibactin exacerbates mismatch repair mutations in colorectal tumors. *Neoplasia* **43**: 100918. <https://www.sciencedirect.com/science/article/pii/S147655862300043X>.
- Dronkert MLG, Kanaar R. 2001. Repair of DNA interstrand cross-links. *Mutation Research/DNA Repair* **486**: 217–247. <https://www.sciencedirect.com/science/article/pii/S0921877701000921>.
- Dubinsky V, Dotan I, Gophna U. 2020. Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. *Trends Microbiol* **28**: 874–876. <https://doi.org/10.1016/j.tim.2020.05.015>.
- Dubois D, Baron O, Cougnoux A, Delmas J, Pradel N, Boury M, Bouchon B, Bringer M-A, Nougayrède J-P, Oswald E, et al. 2011. ClbP Is a Prototype of a Peptidase Subgroup Involved in Biosynthesis of Nonribosomal Peptides. *Journal of Biological Chemistry* **286**: 35562–35570. <https://www.sciencedirect.com/science/article/pii/S0021925820737622>.
- Dubois D, Delmas J, Cady A, Robin F, Sivignon A, Oswald E, Bonnet R. 2010. Cyclomodulins in Urosepsis Strains of Escherichia coli. *J Clin Microbiol* **48**: 2122–2129. <https://doi.org/10.1128/jcm.02365-09>.

- Dziubańska-Kusibab PJ, Berger H, Battistini F, Bouwman BAM, Iftekhar A, Katainen R, Cajuso T, Crosetto N, Orozco M, Aaltonen LA, et al. 2020. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med* **26**: 1063–1069. <https://doi.org/10.1038/s41591-020-0908-2>.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the Human Intestinal Microbial Flora. *Science (1979)* **308**: 1635–1638. <https://doi.org/10.1126/science.1110591>.
- Eklöf V, Löfgren-Burström A, Zingmark C, Edin S, Larsson P, Karling P, Alexeyev O, Rutegård J, Wikberg ML, Palmqvist R. 2017. Cancer-associated fecal microbial markers in colorectal cancer detection. *Int J Cancer* **141**: 2528–2536. <https://pubmed.ncbi.nlm.nih.gov/28833079>.
- Fábrega M-J, Rodríguez-Nogales A, Garrido-Mesa J, Algieri F, Badía J, Giménez R, Gálvez J, Baldomà L. 2017. Intestinal Anti-inflammatory Effects of Outer Membrane Vesicles from Escherichia coli Nissle 1917 in DSS-Experimental Colitis in Mice. *Frontiers in Microbiology* **8**: 1274. <https://www.frontiersin.org/article/10.3389/fmicb.2017.01274>.
- Faïs T, Cougnoux A, Dalmaso G, Laurent F, Delmas J, Bonnet R. 2016. Antibiotic Activity of Escherichia coli against Multiresistant Staphylococcus aureus. *Antimicrob Agents Chemother* **60**: 6986–6988. <https://doi.org/10.1128/AAC.00130-16>.
- Faïs T, Delmas J, Barnich N, Bonnet R, Dalmaso G. 2018. Colibactin: More Than a New Bacterial Toxin. *Toxins (Basel)* **10**: 151.
- Fallahi-Sichani M, Becker V, Izar B, Baker GJ, Lin J, Boswell SA, Shah P, Rotem A, Garraway LA, Sorger PK. 2017. Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Mol Syst Biol* **13**: 905. <https://doi.org/10.15252/msb.20166796>.
- Fallahi-Sichani M, Moerke NJ, Niepel M, Zhang T, Gray NS, Sorger PK. 2015. Systematic analysis of BRAFV600E melanomas reveals a role for JNK/c-Jun pathway in adaptive resistance to drug-induced apoptosis. *Mol Syst Biol* **11**: 797. <https://doi.org/10.15252/msb.20145877>.
- Fan Y, Pedersen O. 2021. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* **19**: 55–71. <https://doi.org/10.1038/s41579-020-0433-9>.
- Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, O'Dwyer PJ, Lee RJ, Grippo JF, Nolop K, et al. 2024. Inhibition of Mutated, Activated

- BRAF in Metastatic Melanoma. *New England Journal of Medicine* **363**: 809–819. <https://doi.org/10.1056/NEJMoa1002011>.
- Flavahan WA, Gaskell E, Bernstein BE. 2017. Epigenetic plasticity and the hallmarks of cancer. *Science (1979)* **357**: eaal2380. <https://doi.org/10.1126/science.aal2380>.
- Fousteri M, Mullenders LHF. 2008. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res* **18**: 73–84. <https://doi.org/10.1038/cr.2008.6>.
- Gaab ME, Lozano PO, Ibañez D, Manese KD, Riego FM, Tiongco RE, Albano PM. 2023. A Meta-Analysis on the Association of Colibactin-Producing pks+ *Escherichia coli* with the Development of Colorectal Cancer. *Lab Med* **54**: 75–82. <https://doi.org/10.1093/labmed/lmac072>.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343–345. <https://doi.org/10.1038/nmeth.1318>.
- Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch S V, Knight R. 2018. Current understanding of the human microbiome. *Nat Med* **24**: 392–400. <https://doi.org/10.1038/nm.4517>.
- Glinghammar B, Inoue H, Rafter JJ. 2002. Deoxycholic acid causes DNA damage in colonic cells with subsequent induction of caspases, COX-2 promoter activity and the transcription factors NF- κ B and AP-1. *Carcinogenesis* **23**: 839–845. <https://doi.org/10.1093/carcin/23.5.839>.
- Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**: 45–50. <https://doi.org/10.1038/nature24287>.
- Goodwin AC, Shields CED, Wu S, Huso DL, Wu X, Murray-Stewart TR, Hacker-Prietz A, Rabizadeh S, Woster PM, Sears CL, et al. 2011. Polyamine catabolism contributes to enterotoxigenic *Bacteroides fragilis*-induced colon tumorigenesis. *Proceedings of the National Academy of Sciences* **108**: 15354–15359. <https://doi.org/10.1073/pnas.1010203108>.
- Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>.
- Green P, Ewing B, Miller W, Thomas PJ, Green ED, Program NCS. 2003. Transcription-associated mutational asymmetry in mammalian evolution. *Nat Genet* **33**: 514–517. <https://doi.org/10.1038/ng1103>.

- Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, Dean M, Huang Y, Jia W, Zhou Q, et al. 2013. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet* **45**: 1459–1463. <https://doi.org/10.1038/ng.2798>.
- H DS, Alvaro B, Grietje H, M JA, J FH, E LG. 2007. Reduced Dietary Intake of Carbohydrates by Obese Subjects Results in Decreased Concentrations of Butyrate and Butyrate-Producing Bacteria in Feces. *Appl Environ Microbiol* **73**: 1073–1078. <https://doi.org/10.1128/AEM.02340-06>.
- Haferkamp S, Borst A, Adam C, Becker TM, Motschenbacher S, Windhövel S, Hufnagel AL, Houben R, Meierjohann S. 2013. Vemurafenib Induces Senescence Features in Melanoma Cells. *Journal of Investigative Dermatology* **133**: 1601–1609. <https://www.sciencedirect.com/science/article/pii/S0022202X15362801>.
- Hallam JC, Sandalli S, Floria I, Turner NCA, Tang-Fichaux M, Oswald E, O’Boyle N, Roe AJ. 2023. D-Serine reduces the expression of the cytopathic genotoxin colibactin. *Microbial Cell* **10**: 63–77.
- Harrison F, Paul J, Massey RC, Buckling A. 2008. Interspecific competition and siderophore-mediated cooperation in *Pseudomonas aeruginosa*. *ISME J* **2**: 49–55. <https://doi.org/10.1038/ismej.2007.96>.
- Hazra TK, Das A, Das S, Choudhury S, Kow YW, Roy R. 2007. Oxidative DNA damage repair in mammalian cells: A new perspective. *DNA Repair (Amst)* **6**: 470–480. <https://www.sciencedirect.com/science/article/pii/S156878640600320X>.
- He Z, Gharaibeh RZ, Newsome RC, Pope JL, Dougherty MW, Tomkovich S, Pons B, Mirey G, Vignard J, Hendrixson DR, et al. 2019. *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* **68**: 289. <http://gut.bmj.com/content/68/2/289.abstract>.
- Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**: 15–25. <https://doi.org/10.1038/nrmicro2259>.
- Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, et al. 2010. Disordered Microbial Communities in Asthmatic Airways. *PLoS One* **5**: e8578-. <https://doi.org/10.1371/journal.pone.0008578>.
- Hoefflin R, Lahrmann B, Warsow G, Hübschmann D, Spath C, Walter B, Chen X, Hofer L, Macher-Goepfing S, Tolstov Y, et al. 2016. Spatial niche

- formation but not malignant progression is a driving force for intratumoural heterogeneity. *Nat Commun* **7**: ncomms11845. <https://doi.org/10.1038/ncomms11845>.
- Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, Zhu D, Koya JB, Wei L, Li J, et al. 2022. Microbiota in health and diseases. *Signal Transduct Target Ther* **7**: 135. <https://doi.org/10.1038/s41392-022-00974-4>.
- Hu J, Selby CP, Adar S, Adebali O, Sancar A. 2017. Molecular mechanisms and genomic maps of DNA excision repair in *Escherichia coli* and humans. *Journal of Biological Chemistry* **292**: 15588–15597.
- Huang JC, Svoboda DL, Reardon JT, Sancar A. 1992. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. *Proceedings of the National Academy of Sciences* **89**: 3664–3668. <https://doi.org/10.1073/pnas.89.8.3664>.
- Huang YJ, Nelson CE, Brodie EL, DeSantis TZ, Baek MS, Liu J, Woyke T, Allgaier M, Bristow J, Wiener-Kronish JP, et al. 2011. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *Journal of Allergy and Clinical Immunology* **127**: 372-381.e3. <https://www.sciencedirect.com/science/article/pii/S0091674910017604>.
- Huo X, Juergens S, Zhang X, Rezaei D, Yu C, Strauch ED, Wang J-Y, Cheng E, Meyer F, Wang DH, et al. 2011. Deoxycholic acid causes DNA damage while inducing apoptotic resistance through NF- κ B activation in benign Barrett's epithelial cells. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **301**: G278–G286. <https://doi.org/10.1152/ajpgi.00092.2011>.
- Iftekhhar A, Berger H, Bouznad N, Heuberger J, Boccellato F, Dobrindt U, Hermeking H, Sigal M, Meyer TF. 2021. Genomic aberrations after short-term exposure to colibactin-producing *E. coli* transform primary colon epithelial cells. *Nat Commun* **12**: 1003. <https://doi.org/10.1038/s41467-021-21162-y>.
- Iyadorai T, Mariappan V, Vellasamy KM, Wanyiri JW, Roslani AC, Lee GK, Sears C, Vadivelu J. 2020. Prevalence and association of pks+ *Escherichia coli* with colorectal cancer in patients at the University Malaya Medical Centre, Malaysia. *PLoS One* **15**: e0228217-. <https://doi.org/10.1371/journal.pone.0228217>.
- Jasin M, Rothstein R. 2013. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb Perspect Biol* **5**: a012740–a012740.

- Jawara A, Stephanie H, L SC, Winston T. 2019. Epigenetic Changes Induced by *Bacteroides fragilis* Toxin. *Infect Immun* **87**: 10.1128/iai.00447-18. <https://doi.org/10.1128/iai.00447-18>.
- Jenkins GJS, Cronin J, Alhamdani A, Rawat N, D'Souza F, Thomas T, Eltahir Z, Griffiths AP, Baxter JN. 2008. The bile acid deoxycholic acid has a non-linear dose response for DNA damage and possibly NF- κ B activation in oesophageal cells, with a mechanism of action involving ROS. *Mutagenesis* **23**: 399–405. <https://doi.org/10.1093/mutage/gen029>.
- Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, Ameh S, Sandel D, Liang XS, Mazzilli S, et al. 2019. Commensal Microbiota Promote Lung Cancer Development via $\gamma\delta$ T Cells. *Cell* **176**: 998-1013.e16. <https://doi.org/10.1016/j.cell.2018.12.040>.
- Jones SA, Jorgensen M, Chowdhury FZ, Rodgers R, Hartline J, Leatham MP, Struve C, Krogfelt KA, Cohen PS, Conway T. 2008. Glycogen and Maltose Utilization by *Escherichia coli* O157:H7 in the Mouse Intestine. *Infect Immun* **76**: 2531–2540. <https://doi.org/10.1128/iai.00096-08>.
- Joshi F, Archana G, Desai A. 2006. Siderophore Cross-Utilization Amongst Rhizospheric Bacteria and the Role of Their Differential Affinities for Fe³⁺ on Growth Stimulation Under Iron-Limited Conditions. *Curr Microbiol* **53**: 141–147. <https://doi.org/10.1007/s00284-005-0400-8>.
- Kaistha A, Levine J. 2014. Inflammatory Bowel Disease: The Classic Gastrointestinal Autoimmune Disease. *Curr Probl Pediatr Adolesc Health Care* **44**: 328–334. <https://www.sciencedirect.com/science/article/pii/S1538544214001096>.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Karanam K, Kafri R, Loewer A, Lahav G. 2012. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. *Mol Cell* **47**: 320–329. <https://doi.org/10.1016/j.molcel.2012.05.052>.
- Keeney KM, Finlay BB. 2011. Enteric pathogen exploitation of the microbiota-generated nutrient environment of the gut. *Curr Opin Microbiol* **14**: 92–98. <https://www.sciencedirect.com/science/article/pii/S1369527410001992>.
- Kern L, Abdeen SK, Kolodziejczyk AA, Elinav E. 2021. Commensal inter-bacterial interactions shaping the microbiota. *Curr Opin Microbiol* **63**: 158–171. <https://www.sciencedirect.com/science/article/pii/S1369527421000989>.

- Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Tiao G, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea AD, et al. 2016. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet* **48**: 600–606. <https://doi.org/10.1038/ng.3557>.
- Kim MS, Chang X, Nagpal JK, Yamashita K, Baek JH, Dasgupta S, Wu G, Osada M, Woo J-H, Westra WH, et al. 2008. The N-methyl-D-aspartate receptor type 2A is frequently methylated in human colorectal carcinoma and suppresses cell growth. *Oncogene* **27**: 2045–2054. <https://doi.org/10.1038/sj.onc.1210842>.
- Korenchan DE, Flavell RR. 2019. Spatiotemporal pH Heterogeneity as a Promoter of Cancer Progression and Therapeutic Resistance. *Cancers (Basel)* **11**: 1026.
- Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. 1994. Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev* **58**: 401–465. <https://doi.org/10.1128/mr.58.3.401-465.1994>.
- Krokan HE, Nilsen H, Skorpen F, Otterlei M, Slupphaug G. 2000. Base excision repair of DNA in mammalian cells. *FEBS Lett* **476**: 73–77. [https://doi.org/10.1016/S0014-5793\(00\)01674-4](https://doi.org/10.1016/S0014-5793(00)01674-4).
- Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, Gomez C, Degasperi A, Harris R, Jackson SP, et al. 2019. A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**: 821-836.e16. <https://doi.org/10.1016/j.cell.2019.03.001>.
- Kudo T, Jeknić S, Macklin DN, Akhter S, Hughey JJ, Regot S, Covert MW. 2018. Live-cell measurements of kinase activity in single cells using translocation reporters. *Nat Protoc* **13**: 155–169. <https://doi.org/10.1038/nprot.2017.128>.
- Kuzminov A. 2011. Homologous Recombination—Experimental Systems, Analysis, and Significance. *EcoSal Plus* **4**.
- LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, Naranjo S, Westcott PMK, Duarte FM, Sankar V, et al. 2020. Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* **38**: 212-228.e13. <https://doi.org/10.1016/j.ccell.2020.06.006>.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.

- Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, et al. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**: 541–546. <https://doi.org/10.1038/nature12506>.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences* **109**: E2774–E2783. <https://doi.org/10.1073/pnas.1210309109>.
- Lee KM, Choi KH, Ouellette MM. 2004. Use of exogenous hTERT to immortalize primary human cells. *Cytotechnology* **45**: 33–38. <https://doi.org/10.1007/s10616-004-5123-3>.
- Lee REC, Walker SR, Savery K, Frank DA, Gaudet S. 2014. Fold Change of Nuclear NF- κ B Determines TNF-Induced Transcription in Single Cells. *Mol Cell* **53**: 867–879. <https://doi.org/10.1016/j.molcel.2014.01.026>.
- Li C, Stražar M, Mohamed AMT, Pacheco JA, Walker RL, Lebar T, Zhao S, Lockart J, Dame A, Thurimella K, et al. 2024. Gut microbiome and metabolome profiling in Framingham heart study reveals cholesterol-metabolizing bacteria. *Cell* **187**: 1834-1852.e19. <https://www.sciencedirect.com/science/article/pii/S0092867424003052>.
- Li G-M. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**: 85–98. <https://doi.org/10.1038/cr.2007.115>.
- Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS. 2014. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**: 554. <https://doi.org/10.1186/s13059-014-0554-4>.
- Li X, Heyer W-D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* **18**: 99–113. <https://doi.org/10.1038/cr.2008.1>.
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**: 417–425.
- Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espéli O, Boccard F, Koszul R. 2018. Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**: 771-783.e18. <https://www.sciencedirect.com/science/article/pii/S0092867417315076>.
- Lito P, Pratilas CA, Joseph EW, Tadi M, Halilovic E, Zubrowski M, Huang A, Wong WL, Callahan MK, Merghoub T, et al. 2012. Relief of Profound

- Feedback Inhibition of Mitogenic Signaling by RAF Inhibitors Attenuates Their Activity in BRAFV600E Melanomas. *Cancer Cell* **22**: 668–682. <https://doi.org/10.1016/j.ccr.2012.10.009>.
- Lloyd MC, Cunningham JJ, Bui MM, Gillies RJ, Brown JS, Gatenby RA. 2016. Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces. *Cancer Res* **76**: 3136–3144. <https://doi.org/10.1158/0008-5472.CAN-15-2962>.
- Lopès A, Billard E, Casse AH, Villéger R, Veziat J, Roche G, Carrier G, Sauvanet P, Briat A, Pagès F, et al. 2020. Colibactin-positive *Escherichia coli* induce a procarcinogenic immune environment leading to immunotherapy resistance in colorectal cancer. *Int J Cancer* **146**: 3147–3159. <https://doi.org/10.1002/ijc.32920>.
- Louis P, Hold GL, Flint HJ. 2014. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* **12**: 661–672. <https://doi.org/10.1038/nrmicro3344>.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Lu R, Wu S, Zhang Y, Xia Y, Liu X, Zheng Y, Chen H, Schaefer KL, Zhou Z, Bissonnette M, et al. 2014. Enteric bacterial protein AvrA promotes colonic tumorigenesis and activates colonic beta-catenin signaling pathway. *Oncogenesis* **3**: e105–e105. <https://doi.org/10.1038/oncsis.2014.20>.
- Lucas C, Salesse L, Hoang MHT, Bonnet M, Sauvanet P, Larabi A, Godfraind C, Gagnière J, Pezet D, Rosenstiel P, et al. 2020. Autophagy of Intestinal Epithelial Cells Inhibits Colorectal Carcinogenesis Induced by Colibactin-Producing *Escherichia coli* in ApcMin/+ Mice. *Gastroenterology* **158**: 1373–1388. <https://www.sciencedirect.com/science/article/pii/S0016508520300068>.
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. 2009. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**: 161. <https://doi.org/10.1186/1471-2105-10-161>.
- Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences* **107**: 961–968. <https://doi.org/10.1073/pnas.0912629107>.
- Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, Agdashian D, Terabe M, Berzofsky JA, Fako V, et al. 2018. Gut microbiome-mediated bile acid

- metabolism regulates liver cancer via NKT cells. *Science (1979)* **360**: ean5931. <https://doi.org/10.1126/science.aan5931>.
- Maddamsetti R, Grant NA. 2020. Divergent Evolution of Mutation Rates and Biases in the Long-Term Evolution Experiment with *Escherichia coli*. *Genome Biol Evol* **12**: 1591–1603. <https://doi.org/10.1093/gbe/evaa178>.
- Mair B, Tomic J, Masud SN, Tonge P, Weiss A, Usaj M, Tong AHY, Kwan JJ, Brown KR, Titus E, et al. 2019. Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Rep* **27**: 599–615.e12.
- Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, Hartmann GG, Smith OC, Kim JY, Evans KV, et al. 2020. Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell* **38**: 229–246.e13. <https://doi.org/10.1016/j.ccell.2020.06.012>.
- Marquet P, Duncan SH, Chassard C, Bernalier-Donadille A, Flint HJ. 2009. Lactate has the potential to promote hydrogen sulphide formation in the human colon. *FEMS Microbiol Lett* **299**: 128–134. <https://doi.org/10.1111/j.1574-6968.2009.01750.x>.
- Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* **15**: 465–481. <https://doi.org/10.1038/nrm3822>.
- Martin P, Marcq I, Magistro G, Penary M, Garcie C, Payros D, Boury M, Olier M, Nougayrède J-P, Audebert M, et al. 2013. Interplay between Siderophores and Colibactin Genotoxin Biosynthetic Pathways in *Escherichia coli*. *PLoS Pathog* **9**: e1003437-. <https://doi.org/10.1371/journal.ppat.1003437>.
- Maslowska KH, Makiela-Dzubska K, Fijalkowska IJ. 2019. The SOS system: A complex and tightly regulated response to DNA damage. *Environ Mol Mutagen* **60**: 368–384. <https://doi.org/10.1002/em.22267>.
- Matsumoto Y, Kim K, Hurwitz J, Gary R, Levin DS, Tomkinson AE, Park MS. 1999. Reconstitution of Proliferating Cell Nuclear Antigen-dependent Repair of Apurinic/Apyrimidinic Sites with Purified Human Proteins*. *Journal of Biological Chemistry* **274**: 33703–33708. <https://www.sciencedirect.com/science/article/pii/S0021925817465197>.
- McGuckin MA, Eri R, Simms LA, Florin THJ, Radford-Smith G. 2009. Intestinal Barrier Dysfunction in Inflammatory Bowel Diseases. *Inflamm Bowel Dis* **15**: 100–113. <https://doi.org/10.1002/ibd.20539>.

- McHugh PJ, Spanswick VJ, Hartley JA. 2001. Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol* **2**: 483–490. [https://doi.org/10.1016/S1470-2045\(01\)00454-5](https://doi.org/10.1016/S1470-2045(01)00454-5).
- McNabney S, Henagan T. 2017. Short Chain Fatty Acids in the Colon and Peripheral Tissues: A Focus on Butyrate, Colon Cancer, Obesity and Insulin Resistance. *Nutrients* **9**: 1348.
- Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, et al. 2012. A framework for human microbiome research. *Nature* **486**: 215–221. <https://doi.org/10.1038/nature11209>.
- Mitchell A, Wei P, Lim WA. 2015. Oscillatory stress stimulation uncovers an Achilles' heel of the yeast MAPK signaling network. *Science (1979)* **350**: 1379–1383. <https://doi.org/10.1126/science.aab0892>.
- Modrich P. 1997. Strand-specific Mismatch Repair in Mammalian Cells. *Journal of Biological Chemistry* **272**: 24727–24730. <https://doi.org/10.1074/jbc.272.40.24727>.
- Moldovan G-L, D'Andrea AD. 2009. How the Fanconi Anemia Pathway Guards the Genome. *Annu Rev Genet* **43**: 223–249.
- Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, et al. 2003. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**: 267–273. <https://doi.org/10.1038/ng1180>.
- Muniandy PA, Liu J, Majumdar A, Liu S, Seidman MM. 2010. DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol* **45**: 23–49. <https://www.tandfonline.com/doi/abs/10.3109/10409230903501819>.
- Nagavardhini A, Chittibabu G. 2022. Meta-Analysis of Altered Gut Microbiota Reveals Microbial and Metabolic Biomarkers for Colorectal Cancer. *Microbiol Spectr* **10**: e00013-22. <https://doi.org/10.1128/spectrum.00013-22>.
- Nakanishi K, Yang Y-G, Pierce AJ, Taniguchi T, Digweed M, D'Andrea AD, Wang Z-Q, Jasin M. 2005. Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proceedings of the National Academy of Sciences* **102**: 1110–1115. <https://doi.org/10.1073/pnas.0407796102>.

- Nakken S, Rødland EA, Hovig E. 2010. Impact of DNA physical properties on local sequence bias of human mutation. *Hum Mutat* **31**: 1316–1325. <https://doi.org/10.1002/humu.21371>.
- Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, et al. 2020. The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science (1979)* **368**: 973 LP – 980. <http://science.sciencemag.org/content/368/6494/973.abstract>.
- Norouzi-Beirami MH, Marashi S-A, Banaei-Moghaddam AM, Kavousi K. 2020. Beyond Taxonomic Analysis of Microbiomes: A Functional Approach for Revisiting Microbiome Changes in Colorectal Cancer. *Front Microbiol* **10**. <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2019.03117>.
- Noto Guillen M, Li C, Rosener B, Mitchell A. 2024. Antibacterial activity of nonantibiotics is orthogonal to standard antibiotics. *Science (1979)* **384**: 93–100. <https://doi.org/10.1126/science.adk7368>.
- Noto Guillen M, Rosener B, Sayin S, Mitchell A. 2021. Assembling stable syntrophic *Escherichia coli* communities by comprehensively identifying beneficiaries of secreted goods. *Cell Syst* **12**: 1064-1078.e7. <https://doi.org/10.1016/j.cels.2021.08.002>.
- Nougayrède J-P, Chagneau C, Motta J-P, Bossuet-Greif N, Belloy M, Taieb F, Gratadoux J-J, Thomas M, Langella P, Oswald E. 2021. A toxic friend: Genotoxic and mutagenic activity of the probiotic strain *Escherichia coli* Nissle 1917. *bioRxiv* 2021.03.22.436450. <http://biorxiv.org/content/early/2021/03/22/2021.03.22.436450.abstract>.
- Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**: 848–851.
- Oliero M, Calvé A, Fragoso G, Cuisiniere T, Hajjar R, Dobrindt U, Santos MM. 2021. Oligosaccharides increase the genotoxic effect of colibactin produced by pks+ *Escherichia coli* strains. *BMC Cancer* **21**: 172. <https://doi.org/10.1186/s12885-021-07876-8>.
- Olivieri M, Cho T, Álvarez-Quilón A, Li K, Schellenberg MJ, Zimmermann M, Hustedt N, Rossi SE, Adam S, Melo H, et al. 2020. A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **182**: 481-496.e21.
- Pagès V, Koffel-Schwartz N, Fuchs RPP. 2003. recX, a new SOS gene that is co-transcribed with the recA gene in *Escherichia coli*. *DNA Repair (Amst)* **2**:

- 273–284.
<https://www.sciencedirect.com/science/article/pii/S1568786402002173>.
- Pfeifer GP, You Y-H, Besaratinia A. 2005. Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**: 19–31.
<https://www.sciencedirect.com/science/article/pii/S0027510704004804>.
- Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmaso G, Stege PB, et al. 2020. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature* **580**: 269–273. <https://doi.org/10.1038/s41586-020-2080-8>.
- Ponder RG, Fonville NC, Rosenberg SM. 2005. A Switch from High-Fidelity to Error-Prone DNA Double-Strand Break Repair Underlies Stress-Induced Mutation. *Mol Cell* **19**: 791–804.
<https://doi.org/10.1016/j.molcel.2005.07.025>.
- Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolk T, Janssen S, Metcalf J, Song SJ, et al. 2020. RETRACTED ARTICLE: Microbiome analyses of blood and tissues suggest cancerdiagnostic approach. *Nature* **579**: 567–574. <https://doi.org/10.1038/s41586-020-2095-1>.
- Poulos RC, Olivier J, Wong JWH. 2017. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. *Nucleic Acids Res* **45**: 7786–7795.
<https://doi.org/10.1093/nar/gkx463>.
- Preibisch S, Saalfeld S, Schindelin J, Tomancak P. 2010. Software for bead-based registration of selective plane illumination microscopy data. *Nat Methods* **7**: 418–419. <https://doi.org/10.1038/nmeth0610-418>.
- Puig I, Tenbaum SP, Chicote I, Arqués O, Martínez-Quintanilla J, Cuesta-Borrás E, Ramírez L, Gonzalo P, Soto A, Aguilar S, et al. 2018. TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. *J Clin Invest* **128**: 3887–3905. <https://doi.org/10.1172/JCI96393>.
- Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, Lahav G. 2012. p53 Dynamics Control Cell Fate. *Science (1979)* **336**: 1440–1444.
<https://doi.org/10.1126/science.1218351>.
- Putnam CD. 2021. Strand discrimination in DNA mismatch repair. *DNA Repair (Amst)* **105**: 103161.

- Regot S, Hughey JJ, Bajar BT, Carrasco S, Covert MW. 2014. High-Sensitivity Measurements of Multiple Kinase Activities in Live Single Cells. *Cell* **157**: 1724–1734. <https://doi.org/10.1016/j.cell.2014.04.039>.
- Reuter C, Alzheimer M, Walles H, Oelschlaeger TA. 2018. An adherent mucus layer attenuates the genotoxic effect of colibactin. *Cell Microbiol* **20**: e12812. <https://doi.org/10.1111/cmi.12812>.
- Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139 *. *Journal of Biological Chemistry* **273**: 5858–5868. <https://doi.org/10.1074/jbc.273.10.5858>.
- Rosendahl Huber A, Pleguezuelos-Manzano C, Puschhof J, Ubels J, Boot C, Saftien A, Verheul M, Trabut LT, Groenen N, van Roosmalen M, et al. 2024. Improved detection of colibactin-induced mutations by genotoxic E. coli in organoids and colorectal cancer. *Cancer Cell* **42**: 487-496.e6. <https://doi.org/10.1016/j.ccell.2024.02.009>.
- Rosener B, Sayin S, Oluoch PO, García González AP, Mori H, Walhout AJM, Mitchell A. 2020. Evolved bacterial resistance against fluoropyrimidines can lower chemotherapy impact in the Caenorhabditis elegans host eds. C.R. Landry, G. Storz, and P. Turnbaugh. *Elife* **9**: e59831. <https://doi.org/10.7554/eLife.59831>.
- Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. 2013. Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/ β -Catenin Signaling via its FadA Adhesin. *Cell Host Microbe* **14**: 195–206. <https://doi.org/10.1016/j.chom.2013.07.012>.
- Salerno P, De Falco V, Tamburrino A, Nappi TC, Vecchio G, Schweppe RE, Bollag G, Santoro M, Salvatore G. 2010. Cytostatic Activity of Adenosine Triphosphate-Competitive Kinase Inhibitors in BRAF Mutant Thyroid Carcinoma Cells. *J Clin Endocrinol Metab* **95**: 450–455. <https://doi.org/10.1210/jc.2009-0373>.
- Salgado H, Gama-Castro S, Lara P, Mejia-Almonte C, Alarcón-Carranza G, López-Almazo AG, Betancourt-Figueroa F, Peña-Loredo P, Alquicira-Hernández S, Ledezma-Tejeida D, et al. 2023. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in E. coli K-12. *Nucleic Acids Res* gkad1072. <https://doi.org/10.1093/nar/gkad1072>.
- Sancar A, Rupp WD. 1983. A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell* **33**: 249–260. <https://www.sciencedirect.com/science/article/pii/0092867483903549>.

- Sayin S, Rosener B, Li CG, Ho B, Ponomarova O, Ward D V, Walhout AJM, Mitchell A. 2023. Evolved bacterial resistance to the chemotherapy gemcitabine modulates its efficacy in co-cultured cancer cells eds. C.R. Landry, P.J. Turnbaugh, and A. Muir. *Elife* **12**: e83140. <https://doi.org/10.7554/eLife.83140>.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682. <https://doi.org/10.1038/nmeth.2019>.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–675. <https://doi.org/10.1038/nmeth.2089>.
- Schofield MJ, Hsieh P. 2003. DNA Mismatch Repair: Molecular Mechanisms and Biological Function. *Annu Rev Microbiol* **57**: 579–608.
- Schwarcz S, Kovács P, Nyerges P, Ujlaki G, Sipos A, Uray K, Bai P, Mikó E. 2024. The bacterial metabolite, lithocholic acid, has antineoplastic effects in pancreatic adenocarcinoma. *Cell Death Discov* **10**: 248. <https://doi.org/10.1038/s41420-024-02023-1>.
- Secher T, Samba-Louaka A, Oswald E, Nougayrède J-P. 2013. Escherichia coli Producing Colibactin Triggers Premature and Transmissible Senescence in Mammalian Cells. *PLoS One* **8**: e77157-. <https://doi.org/10.1371/journal.pone.0077157>.
- Selby CP, Drapkin R, Reinberg D, Sancar A. 1997. RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic Acids Res* **25**: 787–793. <https://doi.org/10.1093/nar/25.4.787>.
- Selby CP, Sancar A. 1990. Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *Journal of Biological Chemistry* **265**: 21330–21336. <https://www.sciencedirect.com/science/article/pii/S0021925817453646>.
- Sender R, Fuchs S, Milo R. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14**: e1002533-. <https://doi.org/10.1371/journal.pbio.1002533>.
- Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. 2024. The microbiome and human cancer. *Science (1979)* **371**: eabc4552. <https://doi.org/10.1126/science.abc4552>.

- Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M, et al. 2017. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**: 431–435. <https://doi.org/10.1038/nature22794>.
- Sharma S V, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, et al. 2010. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* **141**: 69–80. <https://doi.org/10.1016/j.cell.2010.02.027>.
- Shee C, Gibson JL, Darrow MC, Gonzalez C, Rosenberg SM. 2011. Impact of a stress-inducible switch to mutagenic repair of DNA breaks on mutation in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **108**: 13659–13664. <https://doi.org/10.1073/pnas.1104681108>.
- Shraga A, Olshvang E, Davidzohn N, Khoshkenar P, Germain N, Shurrush K, Carvalho S, Avram L, Albeck S, Unger T, et al. 2019. Covalent Docking Identifies a Potent and Selective MKK7 Inhibitor. *Cell Chem Biol* **26**: 98–108.e5. <https://doi.org/10.1016/j.chembiol.2018.10.011>.
- Silpe JE, Wong JWH, Owen S V, Baym M, Balskus EP. 2022. The bacterial toxin colibactin triggers prophage induction. *Nature* **603**: 315–320. <https://doi.org/10.1038/s41586-022-04444-3>.
- Sogari A, Rovera E, Grasso G, Mariella E, Reilly NM, Lamba S, Mauri G, Durinikova E, Vitiello PP, Lorenzato A, et al. 2024. Tolerance to colibactin correlates with homologous recombination proficiency and resistance to irinotecan in colorectal cancer cells. *Cell Rep Med* **5**. <https://doi.org/10.1016/j.xcrm.2023.101376>.
- Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. 2009. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**: 428–432. <https://doi.org/10.1038/nature08012>.
- Stinson BM, Loparo JJ. 2021. Repair of DNA Double-Strand Breaks by the Nonhomologous End Joining Pathway. *Annu Rev Biochem* **90**: 137–164.
- Stohl EA, Brockman JP, Burkle KL, Morimatsu K, Kowalczykowski SC, Seifert HS. 2003. *Escherichia coli* RecX Inhibits RecA Recombinase and Coprotease Activities in Vitro and in Vivo. *Journal of Biological Chemistry* **278**: 2278–2285. <https://doi.org/10.1074/jbc.M210496200>.
- Strakova N, Korena K, Karpiskova R. 2021. *Klebsiella pneumoniae* producing bacterial toxin colibactin as a risk of colorectal cancer development - A systematic review. *Toxicon* **197**: 126–135. <https://www.sciencedirect.com/science/article/pii/S0041010121001112>.

- Straussman R, Morikawa T, Shee K, Barzily-Rokni M, Qian ZR, Du J, Davis A, Mongare MM, Gould J, Frederick DT, et al. 2012. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **487**: 500–504. <https://doi.org/10.1038/nature11183>.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**: 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
- Sun Y, McCorvie TJ, Yates LA, Zhang X. 2020. Structural basis of homologous recombination. *Cellular and Molecular Life Sciences* **77**: 3–18. <https://doi.org/10.1007/s00018-019-03365-1>.
- Supek F, Lehner B. 2017. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* **170**: 534-547.e23. <https://doi.org/10.1016/j.cell.2017.07.003>.
- Thakur BK, Malaisé Y, Martin A. 2019. Unveiling the Mutational Mechanism of the Bacterial Genotoxin Colibactin in Colorectal Cancer. *Mol Cell* **74**: 227–229. <https://www.sciencedirect.com/science/article/pii/S1097276519302771>.
- The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29. <https://doi.org/10.1038/75556>.
- Thompson KJ, Ingle JN, Tang X, Chia N, Jeraldo PR, Walther-Antonio MR, Kandimalla KK, Johnson S, Yao JZ, Harrington SC, et al. 2017. A comprehensive analysis of breast cancer microbiota and host gene expression. *PLoS One* **12**: e0188873-. <https://doi.org/10.1371/journal.pone.0188873>.
- Todd PA, Glickman BW. 1982. Mutational specificity of UV light in *Escherichia coli*: indications for a role of DNA secondary structure. *Proceedings of the National Academy of Sciences* **79**: 4123–4127. <https://doi.org/10.1073/pnas.79.13.4123>.
- Tremblay W, Mompert F, Lopez E, Quaranta M, Bergoglio V, Hashim S, Bonnet D, Alric L, Mas E, Trouche D, et al. 2021. Cytotoxic Distending Toxin Promotes Replicative Stress Leading to Genetic Instability Transmitted to Daughter Cells. *Front Cell Dev Biol* **9**. <https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2021.656795>.

- Tripathi P, Shine EE, Healy AR, Kim CS, Herzon SB, Bruner SD, Crawford JM. 2017. ClbS Is a Cyclopropane Hydrolase That Confers Colibactin Resistance. *J Am Chem Soc* **139**: 17719–17722. <https://pubmed.ncbi.nlm.nih.gov/29112397>.
- Tronnet S, Floch P, Lucarelli L, Gaillard D, Martin P, Serino M, Oswald E. 2020. The Genotoxin Colibactin Shapes Gut Microbiota in Mice. *mSphere* **5**: 10.1128/msphere.00589-20. <https://doi.org/10.1128/msphere.00589-20>.
- Tronnet S, Garcie C, Brachmann AO, Piel J, Oswald E, Martin P. 2017. High iron supply inhibits the synthesis of the genotoxin colibactin by pathogenic *Escherichia coli* through a non-canonical Fur/RyhB-mediated pathway. *Pathog Dis* **75**: ftx066. <https://doi.org/10.1093/femspd/ftx066>.
- Tronnet S, Garcie C, Rehm N, Dobrindt U, Oswald E, Martin P. 2016. Iron Homeostasis Regulates the Genotoxicity of *Escherichia coli* That Produces Colibactin. *Infect Immun* **84**: 3358–3368. <https://doi.org/10.1128/iai.00659-16>.
- Tsabar M, Lovitch SB, Jambhekar A, Lahav G. 2021. Connecting Timescales in Biology: Can Early Dynamical Measurements Predict Long-Term Outcomes? *Trends Cancer* **7**: 301–308. <https://doi.org/10.1016/j.trecan.2020.12.008>.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al. 2009. A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484. <https://doi.org/10.1038/nature07540>.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. *Nature* **449**: 804–810. <https://doi.org/10.1038/nature06244>.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031. <https://doi.org/10.1038/nature05414>.
- Vasquez KM, Marburger K, Intody Z, Wilson JH. 2001. Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences* **98**: 8403–8410. <https://doi.org/10.1073/pnas.111009698>.
- Velilla JA, Volpe MR, Kenney GE, Walsh RM, Balskus EP, Gaudet R. 2023. Structural basis of colibactin activation by the ClbP peptidase. *Nat Chem Biol* **19**: 151–158. <https://doi.org/10.1038/s41589-022-01142-z>.

- Vermeulen W, Fousteri M. 2013. Mammalian Transcription-Coupled Excision Repair. *Cold Spring Harb Perspect Biol* **5**: a012625–a012625.
- Veziat J, Villéger R, Barnich N, Bonnet M. 2021. Gut Microbiota as Potential Biomarker and/or Therapeutic Target to Improve the Management of Cancer: Focus on Colibactin-Producing *Escherichia coli* in Colorectal Cancer. *Cancers* **13**.
- Vizcaino MI, Crawford JM. 2015. The colibactin warhead crosslinks DNA. *Nat Chem* **7**: 411–417. <https://doi.org/10.1038/nchem.2221>.
- Vogel SM, Bauer MR, Joerger AC, Wilcken R, Brandt T, Veprintsev DB, Rutherford TJ, Fersht AR, Boeckler FM. 2012. Lithocholic acid is an endogenous inhibitor of MDM4 and MDM2. *Proceedings of the National Academy of Sciences* **109**: 16906–16910. <https://doi.org/10.1073/pnas.1215060109>.
- Vollmer AC, Belkin S, Smulski DR, K VDT, LaRossa RA. 1997. Detection of DNA damage by use of *Escherichia coli* carrying *recA'*::lux, *uvrA'*::lux, or *alkA'*::lux reporter plasmids. *Appl Environ Microbiol* **63**: 2566–2571. <https://doi.org/10.1128/aem.63.7.2566-2571.1997>.
- Volpe MR, Velilla JA, Daniel-Ivad M, Yao JJ, Stornetta A, Villalta PW, Huang H-C, Bachovchin DA, Balbo S, Gaudet R, et al. 2023. A small molecule inhibitor prevents gut bacterial genotoxin production. *Nat Chem Biol* **19**: 159–167. <https://doi.org/10.1038/s41589-022-01147-8>.
- Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, Brown D, Stares MD, Scott P, Bergerat A, et al. 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**: 220–230. <https://doi.org/10.1038/ismej.2010.118>.
- Wami H, Wallenstein A, Sauer D, Stoll M, von Büнау R, Oswald E, Müller R, Dobrindt U. 2021. Insights into evolution and coexistence of the colibactin- and yersiniabactin secondary metabolite determinants in enterobacterial populations. *Microb Genom* **7**. <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000577>.
- Wandersman C, Delepelaire P. 2004. Bacterial Iron Sources: From Siderophores to Hemophores. *Annu Rev Microbiol* **58**: 611–647.
- Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**: 147. <https://doi.org/10.1038/ncomms1146>.

- Watanabe D, Murakami H, Ohno H, Tanisawa K, Konishi K, Tsunematsu Y, Sato M, Miyoshi N, Wakabayashi K, Watanabe K, et al. 2020. Association between dietary intake and the prevalence of tumourigenic bacteria in the gut microbiota of middle-aged Japanese adults. *Sci Rep* **10**: 15221. <https://doi.org/10.1038/s41598-020-72245-7>.
- Waters LS, Minesinger BK, Wilttrout ME, D'Souza S, Woodruff R V., Walker GC. 2009. Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiology and Molecular Biology Reviews* **73**: 134–154.
- Wernke KM, Tirla A, Xue M, Surovtseva Y V, Menges FS, Herzon SB. 2021. Probing Microbiome Genotoxicity: A Stable Colibactin Provides Insight into Structure–Activity Relationships and Facilitates Mechanism of Action Studies. *J Am Chem Soc* **143**: 15824–15833. <https://doi.org/10.1021/jacs.1c07559>.
- Wernke KM, Xue M, Tirla A, Kim CS, Crawford JM, Herzon SB. 2020. Structure and bioactivity of colibactin. *Bioorg Med Chem Lett* **30**: 127280. <https://www.sciencedirect.com/science/article/pii/S0960894X20303905>.
- West SA, Buckling A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc R Soc Lond B Biol Sci* **270**: 37–44.
- Weterings E, Chen DJ. 2008. The endless tale of non-homologous end-joining. *Cell Res* **18**: 114–124. <https://doi.org/10.1038/cr.2008.3>.
- Wilson MR, Jiang Y, Villalta PW, Stornetta A, Boudreau PD, Carrá A, Brennan CA, Chun E, Ngo L, Samson LD, et al. 2019. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**: eaar7785. <https://pubmed.ncbi.nlm.nih.gov/30765538>.
- Wong JJ, Ho FK, Choo PY, Chong KKL, Ho CMB, Neelakandan R, Keogh D, Barkham T, Chen J, Liu CF, et al. 2022. Escherichia coli BarA-UvrY regulates the pks island and kills Staphylococci via the genotoxin colibactin during interspecies competition. *PLoS Pathog* **18**: e1010766-. <https://doi.org/10.1371/journal.ppat.1010766>.
- Wood RD. 2010. Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environ Mol Mutagen* **51**: 520–526.
- Wood RD. 1997. Nucleotide Excision Repair in Mammalian Cells. *Journal of Biological Chemistry* **272**: 23465–23468.

- Wood TE, Howard SA, Förster A, Nolan LM, Manoli E, Bullen NP, Yau HCL, Hachani A, Hayward RD, Whitney JC, et al. 2019. The *Pseudomonas aeruginosa* T6SS Delivers a Periplasmic Toxin that Disrupts Bacterial Cell Morphology. *Cell Rep* **29**: 187-201.e7.
<https://doi.org/10.1016/j.celrep.2019.08.094>.
- Wright WD, Shah SS, Heyer W-D. 2018. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry* **293**: 10524–10535. <https://doi.org/10.1074/jbc.TM118.000372>.
- Xue M, Kim CS, Healy AR, Wernke KM, Wang Z, Frischling MC, Shine EE, Wang W, Herzon SB, Crawford JM. 2019. Structure elucidation of colibactin and its DNA cross-links. *Science* **365**: eaax2685.
<https://pubmed.ncbi.nlm.nih.gov/31395743>.
- Xue M, Shine E, Wang W, Crawford JM, Herzon SB. 2018. Characterization of Natural Colibactin–Nucleobase Adducts by Tandem Mass Spectrometry and Isotopic Labeling. Support for DNA Alkylation by Cyclopropane Ring Opening. *Biochemistry* **57**: 6391–6394.
<https://doi.org/10.1021/acs.biochem.8b01023>.
- Zhai Y, Minnick PJ, Pribis JP, Garcia-Villada L, Hastings PJ, Herman C, Rosenberg SM. 2023. ppGpp and RNA-polymerase backtracking guide antibiotic-induced mutable gambler cells. *Mol Cell* **83**: 1298-1310.e4.
<https://www.sciencedirect.com/science/article/pii/S1097276523001600>.